

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2002, 14:59:10 ; Search time 24.84 seconds

(without alignments)  
377.193 Million cell updates/sec

Title: US-09-591-632-2\_COPY\_1\_123

Perfect score: 701

Sequence: 1 MSDSNOGNNQNTQYQSQNG.....FNVNNLQGYQAGFQPSQG 123

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	701	100.0	685	1	EFBY52
2	166	23.7	405	2	S19355
3	166	23.7	741	2	S12921
4	163.5	23.3	589	2	S71954
5	148	21.1	419	2	T04886
6	147.5	21.0	409	2	T20847
7	147.5	21.0	945	2	T21998
8	147	21.0	462	1	LIND07
9	144	20.5	232	2	T21735
10	143	20.4	372	2	T64223
11	142	20.3	676	2	S41022
12	141	20.1	404	2	S46269
13	141	20.1	800	2	T51653
14	140.5	20.0	236	2	T18240
15	139	19.8	632	2	T02627
16	138	19.7	1111	2	T28070
17	137.5	19.6	284	2	T23158
18	137.5	19.6	755	2	T15447
19	137	19.5	461	2	C29349
20	136.5	19.5	461	2	T10265
21	136	19.4	462	4	S33798
22	136	19.4	526	1	S33799
23	135	19.3	128	2	S45871
24	135	19.3	528	2	G02127
25	135	19.3	867	2	T27136
26	135	19.3	871	2	T27135
27	133	19.0	3498	2	T22330
28	132	18.8	882	2	H70199
29	131	18.7	1161	2	T18400

## ALIGNMENTS

30	128	18.3	618	2	T19407	hypothetical prote
31	128	18.3	735	2	T19403	hypothetical prote
32	126.5	18.0	738	2	S37876	glutamine-rich pro
33	124.5	17.8	382	2	T29339	hypothetical prote
34	124.5	17.8	1196	2	S65245	translation elonga
35	123	17.5	346	2	S44874	ZC21.3 protein - C
36	121.5	17.3	353	1	S56750	single stranded D
37	121	17.3	597	2	T51889	related to clathri
38	121	17.3	954	1	S20907	endo-1,4-beta-xyla
39	120.5	17.2	989	2	T02568	hypothetical prote
40	119.5	17.0	301	2	JM0079	heterogeneous nucl
41	119.5	17.0	557	2	T19686	hypothetical prote
42	119.5	17.0	1139	2	T33275	hypothetical prote
43	118.5	16.9	129	2	T06256	dormancy-associate
44	118	16.8	592	2	T21536	hypothetical prote
45	118	16.8	838	1	EWTHW	glutennin, high mol

## RESULT 1

EFBY52  
Suppressor 2 protein - yeast (Saccharomyces cerevisiae)  
N:Alternate names: G1-to-S transition protein; protein YD9395.05; protein YDR172W  
C:Species: Saccharomyces cerevisiae  
C:Date: 31-Dec-1991 #sequence revision 31-Dec-1991 #extl change 19-Jan-2001  
C:Accession: S00733; J030323; S49768; S00488; A26742; S00533; S05723  
R:Wilson, P.G.; Gilbertson, M.R.  
J. Mol. Biol. 199, 559-573, 1988  
A:Title: SUP2 suppressor protein of yeast. A fusion protein related to the EF-1 faml  
A:Reference number: S00733; MUID:88172503  
A:Accession: S00733  
A:Molecule type: DNA  
A:Residues: 1-685 <MIL>  
A:Cross-references: EMBL:X07163; NID:94581; PIDN:CAA30155.1; PID:94582  
R:Kushnir, V.V.; Ter-Avanesyan, M.D.; Telickov, M.V.; Surguchov, A.P.; Smirnov, V.N.  
Gene 66, 45-54, 1988  
A:Title: Nucleotide sequence of the SUP2(SUP35) gene of Saccharomyces cerevisiae.  
A:Reference number: J030323; MUID:88329727  
A:Accession: J030323  
A:Molecule type: DNA  
A:Residues: 1-685 <KUS>  
A:Cross-references: EMBL:M21129; NID:g172789; PIDN:AAA313.1; PID:g172791  
R:Murphy, L.; Harris, D.E.  
Submitted to the EMBL Data Library, November 1994  
A:Reference number: S49764  
A:Accession: S49768  
A:Molecule type: DNA  
A:Residues: 1-685 <MUR>  
A:Cross-references: EMBL:Z46727; NID:g1289283; PIDN:CAA6677.1; PID:g1289287; GSPDB:G  
R:Kikuchi, Y.; Shimatake, H.; Kikuchi, A.  
EMBO J. 7, 1175-1182, 1988  
A:Title: A yeast gene regulated for the G1-to-S transition encodes a protein containin  
A:Reference number: S00488; MUID:88296422  
A:Accession: S00488  
A:Molecule type: DNA  
A:Residues: 1-52, 'C', 54-685 <KIK>  
A:Cross-references: GB:Y00829; EMBL:Y00859; NID:g3711; PIDN:CAA68760.1; PID:g3712  
C:Genetics:  
A:Gene: SCD: SUP35; SUP12; GST1; SUP2; MIPS:YDR172W  
A:Cross-references: SCD: S0002579; MIPS:YDR172W  
A:Map position: 4R  
C:Superfamily: suppressor 2 protein; translation elongation factor Tu homology  
C:Keywords: duplication; GTP binding; nucleotide binding; P-loop; phosphoprotein; tan  
F:1-123/Domain: A <DOM1>  
F:42-119/Region: 10-residue repeats  
F:124-253/Domain: charged <DOM2>  
F:159-222/Region: glutamic acid/lysine-rich  
F:254-685/Domain: C <DOM4>  
F:261-409/Domain: translation elongation factor Tu homology <ETU>  
F:267-274/Region: nucleotide-binding motif A (P-loop)  
F:406-409/Region: GTP-binding NKXD motif



A:Title: Dictyostellium annexin VII (Synexin). cDNA sequence and isolation of a gene d	
A:Reference number: A40977; MUID:913/3375	
A:Accession: A40977	







A:Reference number: Z14681  
 A:Accession: T02627  
 A>Status: translated from GB/EMBL/DBD  
 A:Molecule type: DNA  
 A:Residues: 1-632 <R0U>  
 A:Cross-references: EMBL:AC004747; NID:g3413696; PID:g3413716  
 A:Experimental source: cultivar Columbia  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentto, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
 Naure, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487  
 A:Accession: H84654  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-632 <STO>  
 A:Cross-references: GB:AE02093; NID:g3413716; PID:MAC31239.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: T19L18.23; At2g25960  
 A:Map position: 2  
 A:Introns: 112/1; 206/3; 240/3; 264/3; 304/3  
 C:Superfamily: glutenin

Query Match 19.8%; Score 139; DB 2; Length 632;  
 Best local Similarity 31.7%; Pred. No.5.8e-05;  
 Matches 46; Conservative 6; Mismatches 53; Indels 40; Gaps 7;  
 QY 13 YQYYSQNGNQOGNNRYQGYQAYNAQAQAPAG-----YQNYQGY----SGYQ 56  
 Db 374 YDQSSVPPSQSQAQGEYDYGGQSQSQSSGSSAPPTDTGTGYNYQHASGQAGQGYQ 433  
 QY 57 QGGYQOYNPDAGYQOQYNPQGGYQGYNPQGYQOQGFNP-----QGGRGNY 101  
 Db 434 QDGYGAYN--ASQDSGSGQAAGYDQ---QGGYSTTNPSQEDASQAAPPSSASQSGQAGY 488  
 QY 102 KNFNYNNTLQGY--QAGF--QPOSQ 122  
 Db 489 GTTGQOPRPAQSGTGQAQGYGAPPTSQ 513

Search completed: February 1, 2002, 15:02:47  
 Job time: 217 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2002, 14:58:20 ; Search time 19.74 Seconds  
(without alignments)  
140.218 Million cell updates/sec

Title: US-09-591-632-2\_COPY\_1\_123

Perfect score: 701  
Sequence: 1 MSDSNQNMNQNYQYQSQNG.....FNYNINLQGYQAGFQPOSOG 123

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*\n2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*\n3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*\n4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*\n5: /cgn2\_6/ptodata/2/1aa/PCITUS\_COMB.pep:\*\n6: /cgn2\_6/ptodata/2/1aa/backfillsl.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	17.3	485	2	US-08-749-391-2
2	121	17.3	485	4	US-09-390-200-2
3	111.5	15.9	296	2	US-08-700-637-4
4	104	14.8	307	1	US-07-982-112-2
5	103.5	14.8	362	1	US-08-437-027-21
6	103.5	14.8	365	2	US-08-437-027-20
7	103.5	14.8	656	2	US-08-343-443B-2
8	103.5	14.8	656	4	US-09-214-564A-4
9	103	14.7	341	2	US-08-538-711A-8
10	103	14.7	341	4	US-08-725-027-8
11	103	14.7	353	2	US-08-538-711A-7
12	103	14.7	353	4	US-08-725-027-7
13	93	13.3	542	1	US-07-814-964-13
14	93	13.3	542	1	US-08-258-442-13
15	93	13.3	542	1	US-08-328-809-8
16	93	13.3	542	3	PCIT-US92-11107-13
17	91	13.0	1160	5	US-08-808-599A-24
18	90	12.8	493	4	US-08-556-978B-59
19	90	12.8	529	4	US-09-247-806-2
20	90	12.8	729	4	US-09-625-188-20
21	89.5	12.8	1895	2	US-08-619-554-4
22	89.5	12.8	430	2	US-08-945-848-8
23	89.5	12.8	806	1	US-07-980-528-2
24	89	12.7	303	4	US-08-818-112-92
25	87	12.4	405	2	US-08-282-197C-61
26	86	12.3	678	5	PCIT-US93-03027-3
27	85	12.1	595	1	US-08-425-069-4

28	85	12.1	595	2	US-08-317-844B-4	Sequence 4, Appl1
29	84	12.0	545	2	US-08-990-114-1	Sequence 1, Appl1
30	83.5	11.9	655	4	US-09-347-833-4	Sequence 4, Appl1
31	81.5	11.6	660	3	US-09-058-489-18	Sequence 18, Appl1
32	81.5	11.6	660	3	US-09-058-489-91	Sequence 91, Appl1
33	81.5	11.6	1876	2	US-08-619-554-2	Sequence 2, Appl1
34	79	11.3	714	4	US-08-556-978B-61	Sequence 61, Appl1
35	79	11.3	714	4	US-09-247-806-10	Sequence 10, Appl1
36	78.5	11.2	724	3	US-08-793-331-4	Sequence 6, Appl1
37	78.5	11.2	725	3	US-08-793-331-6	Sequence 6, Appl1
38	77.5	11.1	314	2	US-08-525-742-6	Sequence 6, Appl1
39	77.5	11.1	788	2	US-08-918-914-4	Sequence 4, Appl1
40	76.5	10.9	234	4	US-09-146-770-4	Sequence 4, Appl1
41	76.5	10.9	234	4	US-09-216-295-4	Sequence 4, Appl1
42	76	10.8	227	1	US-08-458-023B-4	Sequence 4, Appl1
43	76	10.8	396	1	US-07-702-771-1	Sequence 1, Appl1
44	76	10.8	396	1	US-08-438-883-1	Sequence 1, Appl1
45	76	10.8	396	1	US-08-741-599-1	Sequence 1, Appl1

## ALIGNMENTS

RESULT 1

US-08-749-391-2

Patent No. 5948667

GENERAL INFORMATION:

APPLICANT: Cheng, Kuo-Joan

APPLICANT: Sellinger, Leonard B.

APPLICANT: Liu, Jin-Hao

APPLICANT: Hu, Youji

APPLICANT: Forsberg, Cecil W.

APPLICANT: Moloney, Maurice W.

TITLE OF INVENTION: A Xylanase Obtained From an

TITLE OF INVENTION: Anaerobic Fungus

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Greenlee, Winner and Sullivan, P.C.

STREET: 5370 Manhattan Circle, Suite 201

CITY: Boulder

STATE: Colorado

COUNTRY: USA

ZIP: 80803

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US-08-749,391

FILING DATE: 13-NOV-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Donna M. Feiber

REGISTRATION NUMBER: 33,878

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 499-8080

TELEFAX: (303) 499-8089

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 485 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-749-391-2

Query Match 17.38; Score 121; DB 2; Length 485;  
Best Local Similarity 35.5%; Pred. No. 0.00013;  
Matches 49; Conservative 8; Mismatches 55; Indels 26; Gaps 8;



NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nikaido, Marmelstein, Murray & Oram  
STREET: 655 Fifteenth Street N.W. Suite 330  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-5701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/982,112  
FILING DATE: 19921125  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/617,102  
FILING DATE: 21-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Chin, Monica F.  
REGISTRATION NUMBER: 36,105  
REFERENCE/DOCKET NUMBER: 564-1906  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)638-5000  
TELEFAX: (202)638-4810  
INFORMATION FOR SEQ. ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 307 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-982-112-2

Query Match 14.8%; Score 104; DB 1; Length 307;  
Best Local Similarity 39.7%; Pred. No. 0.0038;  
Matches 29; Conservative 6; Mismatches 28; Indels 10; Gaps 5;  
QY 38 QADPAGYVNTGYSGYQ--GGR--QQYNPDAGYQOQY-NPQGYQOYNPQGGY-QQ 90  
DB 236 EKQPAPE--DAYGDAGYGGPGGPGQDYGPGGYQPDYGGPAGSAGGPGGQDYQOQ 292  
QY 91 QFNPDGGRGNKYN 103  
DB 293 GYGQGGAPTSFSN 305

RESULT 5  
US-08-437-027-21  
Sequence 21, Application US/08437027  
Patent No. 5670317  
GENERAL INFORMATION:  
APPLICANT: Landany1, Marc  
TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/437,027

FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 46416/JPW/CCA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ. ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 362 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-437-027-21

Query Match 14.8%; Score 103.5; DB 1; Length 362;  
Best Local Similarity 31.8%; Pred. No. 0.0052;  
Matches 35; Conservative 4; Mismatches 48; Indels 23; Gaps 4;  
QY 12 NYQYSGNQNGQGNRYGYQAYNAQAQAPAGYGYQYQYSGYQGGYQOYNPDAGYQO 71  
DB 5 DYSTYSQAAQ-----QGYSAVT--AQPTGYAQTTQAGQGSYGYGQPTDVSYTA 55  
QY 72 QYNPQGYQOY-----NPQGYQOQFNPQGGRGKYNKNNYNNNQGYQAG 116  
DB 56 QTNATYGTATATSYGQPPGTGTTPTAPQA-----TSQPVQGYGTG 96

RESULT 6  
US-08-437-027-20  
Sequence 20, Application US/08437027  
Patent No. 5670317  
GENERAL INFORMATION:  
APPLICANT: Landany1, Marc  
TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/437,027  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 46416/JPW/CCA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ. ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 365 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-437-027-20



CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/538,711A  
FILING DATE: 02-OCT-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: KATHRYN M. BROWN  
REGISTRATION NUMBER: 34,556  
REFERENCE/DOCKET NUMBER: 2026-4201  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 751-6849  
TELEFAX: (212) 751-6849  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 341  
TYPE: Amino Acid  
STRANDEDNESS: Unknown  
TOPOLOGY: Linear  
MOLECULE TYPE: peptide  
US-08-538-711A-8

Query Match 14.7%; Score 103; DB 2; Length 341;  
Best Local Similarity 33.3%; Pred. No. 0.0054;  
Matches 33; Conservative 6; Mismatches 44; Indels 16; Gaps 6;

OY 18 QNGNQOQGNR--YQGYQAYNMAQAPAGGYQYQGYSGYQOQGYOQYNDAGYQOQYNP 75

Db 216 RGSQSGSGSGRGGRGDBYNGY--GGPGGNGFGSGPGYGG--GRGCGGGGPGYG-----NQ 267

OY 76 QGGYQOYNPQGYQOQFNPQGGRGYKNF--NYNNNLQGY 113

Db 268 GGGY-----GGGYDNYGGGNGYSGNINDEGNYNQPSNT 301

RESULT 10

US-08-725-027-8  
Sequence 8, Application US/08725027  
Patent No. 6251586

GENERAL INFORMATION:

APPLICANT: MULSHINE, JAMES, L.

APPLICANT: TOCKMAN, MELVYN, S.

TITLE OF INVENTION: AN EPITHELIAL PROTEIN AND

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN, L.L.P.

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/725,027

FILING DATE: 02-OCT-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US08/538,711

FILING DATE: 02-OCT-1995

ATTORNEY/AGENT INFORMATION:

NAME: KATHRYN M. BROWN

REGISTRATION NUMBER: 34,556

REFERENCE/DOCKET NUMBER: 2026-4201US1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 751-6849

TELEFAX: (212) 751-6849

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:  
LENGTH: 341  
TYPE: Amino Acid  
STRANDEDNESS: Unknown  
TOPOLOGY: Linear  
MOLECULE TYPE: peptide  
US-08-725-027-8

Query Match 14.7%; Score 103; DB 4; Length 341;  
Best Local Similarity 33.3%; Pred. No. 0.0054;  
Matches 33; Conservative 6; Mismatches 44; Indels 16; Gaps 6;

OY 18 QNGNQOQGNR--YQGYQAYNMAQAPAGGYQYQGYSGYQOQGYOQYNDAGYQOQYNP 75

Db 216 RGSQSGSGSGRGGRGDBYNGY--GGPGGNGFGSGPGYGG--GRGCGGGGPGYG-----NQ 267

OY 76 QGGYQOYNPQGYQOQFNPQGGRGYKNF--NYNNNLQGY 113

Db 268 GGGY-----GGGYDNYGGGNGYSGNINDEGNYNQPSNT 301

RESULT 11

US-08-538-711A-7

Sequence 7, Application US/08538711A

Patent No. 5994062

GENERAL INFORMATION:

APPLICANT: MULSHINE, JAMES, L.

TITLE OF INVENTION: AN EPITHELIAL PROTEIN AND

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN, L.L.P.

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/538,711A

FILING DATE: 02-OCT-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: KATHRYN M. BROWN

REGISTRATION NUMBER: 34,556

REFERENCE/DOCKET NUMBER: 2026-4201

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 751-6849

TELEFAX: (212) 751-6849

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 353

TYPE: Amino Acid

STRANDEDNESS: Unknown

TOPOLOGY: Linear

MOLECULE TYPE: peptide

US-08-538-711A-7

Query Match 14.7%; Score 103; DB 2; Length 353;

Best Local Similarity 33.3%; Pred. No. 0.0056;

Matches 33; Conservative 6; Mismatches 44; Indels 16; Gaps 6;

OY 18 QNGNQOQGNR--YQGYQAYNMAQAPAGGYQYQGYSGYQOQGYOQYNDAGYQOQYNP 75

Db 216 RGSQSGSGSGRGGRGDBYNGY--GGPGGNGFGSGPGYGG--GRGCGGGGPGYG-----NQ 267

OY 76 QGGYQOYNPQGYQOQFNPQGGRGYKNF--NYNNNLQGY 113

Db 268 GGGY-----GGGYDNYGGGNGYSGNINDEGNYNQPSNT 301

RESULT 11









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2002, 14:57:35 ; Search time 38.77 Seconds  
(without alignments)  
235.002 Million cell updates/sec

Title: US-09-591-632-2-COPY\_1\_123  
Perfect score: 701  
Sequence: 1 MSDSNOCNNQNYQOYSQNG.....FNNYNNLQGYQAGFQPOSOG 123

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: A.Geneseq\_1101.\*  
2: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
3: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
4: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
5: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
6: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
7: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
8: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
9: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
10: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
11: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
12: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
13: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
14: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
15: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
16: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
17: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
18: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
19: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
20: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
21: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
22: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
23: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	701	100.0	685	22	AAB30792	Amino acid sequenc
2	682	97.3	271	22	AAB30800	A modified N regio
3	450	64.2	215	22	AAB30799	A modified N regio
4	228	32.5	715	22	AAB30820	Amino acid sequenc
5	166	23.7	405	22	AAB30810	Amino acid sequenc
6	166	23.7	405	22	AAB30821	Amino acid sequenc
7	166	23.7	741	22	AAB30819	Amino acid sequenc
8	166	22.8	407	10	AA94879	ORF incorporated w
9	148	21.1	417	21	AA630013	Arabidopsis thalia
10	148	21.1	419	21	AA630012	Arabidopsis thalia
11	148	21.1	439	21	AA630011	Arabidopsis thalia

12	141	20.1	404	17	AAR90675	Human chromosome-1
13	136.5	19.5	461	16	AAR75506	Nicotiana glauca ar
14	135	19.3	128	22	AAB30822	Amino acid sequenc
15	134	19.1	341	11	AAR06522	Tapetum-specific p
16	134	19.1	341	11	AAR09307	Male flower-specif
17	134	19.1	341	11	AAR09298	Male flower-specif
18	134	19.1	341	18	AAW10628	Anther-specific CD
19	132	18.8	437	16	AAR75507	Nicotiana glauca
20	129	18.4	341	11	AAR06451	PM510 contg. male
21	128	18.3	894	22	AAB35148	Human NFAR-2 SEQ I
22	121	17.3	280	21	AAB63197	Gene 21 human secr
23	121	17.3	485	19	AAW65462	Neocallimastix pat
24	121	17.3	485	20	AAV27283	N. patriciarum xyl
25	121	17.3	485	22	AAB35588	N. patriciarum end
26	121	17.3	954	21	AAW85550	Ruminococcus flave
27	120.5	17.2	989	21	AAW33723	Arabidopsis thalia
28	120.5	17.2	994	21	AAW33722	Arabidopsis thalia
29	120.5	17.2	1006	21	AAW33721	Arabidopsis thalia
30	118.5	16.9	398	21	AAW29375	Arabidopsis thalia
31	118.5	16.9	418	21	AAW29374	Arabidopsis thalia
32	118.5	16.9	432	21	AAW29373	Arabidopsis thalia
33	117	16.7	579	22	AAB83003	Mealworm encapsula
34	117	16.7	579	22	AAB83004	Mealworm encapsula
35	116	16.5	545	13	AAW27493	p34 from M. paratu
36	115	16.4	103	21	AAW95064	Candida albicans p
37	115	16.4	171	21	AAW30911	Arabidopsis thalia
38	115	16.4	266	21	AAW30910	Arabidopsis thalia
39	115	16.4	324	21	AAW30909	Arabidopsis thalia
40	114	16.3	103	21	AAW95066	Arabidopsis thalia
41	111.5	15.9	946	21	AAW43595	Candida albicans p
42	111.5	15.9	972	21	AAW43594	Arabidopsis thalia
43	111.5	15.9	1026	21	AAW43593	Arabidopsis thalia
44	111	15.8	256	22	AAB30807	Amino acid sequenc
45	108.5	15.5	128	21	AAB58406	Lung cancer associ

## ALIGNMENTS

RESULT 1	
ID AAB30792	standard; Protein: 685 AA.
XX	
AC AAB30792:	
XX	
DT 02-APR-2001	(first entry)
XX	
DE	Amino acid sequence of a yeast Sup35 protein
XX	
KW	Sup35; phenotype: SCHNG; self-coalesce; higher-order aggregate;
KW	amyloidogenic domain; aggregation; fibril; phenotypic alteration;
KW	gene therapy; disease resistance; plant pigmentation; prion disease.
XX	
OS	Saccharomyces cerevisiae.
PN	W0200075324-A2.
XX	
PD	14-DEC-2000.
XX	
PF	09-JUN-2000; 2000MO-US15876.
XX	
PR	09-JUN-1999; 99US-0138833.
XX	
PA	(ARCH-) ARCH DEV CORP.
XX	
PI	Linquist S, Li L, Ma J, Liu J, Sondheimer N, Scheibel T;
XX	
DR	WPI: 2001-061723/07.
XX	
PT	N-PSDB; AAC6679.
XX	
PT	New nucleic acid encoding chimeric proteins with self-assembly
PT	properties, useful e.g. for diagnosis and treatment of prion diseases,
PT	also related aggregates, fibrils and polymers -

XX Claim 11; Page 125-127; 188bp; English.  
PS  
XX The present sequence represents a yeast Sup35 protein. The protein  
CC possesses the prion-like capacity to undergo a self-perpetuating  
CC conformational alteration that changes the functional state of Sup35 in  
CC a manner that creates a heritable change in phenotype. It is used to  
CC construct chimeric polypeptides of the invention, which comprise at  
CC least one SCHAG (self-coalesces into higher-order aggregates) amino acid  
CC sequence fused in frame with a polypeptide of interest (which is other  
CC than a marker protein, a glutathione-S-transferase or a staphylococcal  
CC nuclear protein). The specification also describes chimeric polypeptides  
CC that comprises an amyloidogenic domain that causes aggregation into  
CC fibrils. The chimeric polypeptides are used to prepare polymers with  
CC multiple reactivities, e.g. derivatised with enzymes, or specific  
CC binding partners, and useful e.g. for performing multi-step chemical  
CC reactions. They can be used create an inducible, or stable phenotypic  
CC alteration in a cell, e.g. for gene therapy, protein production,  
CC imparting disease resistance to plants, altering plant pigmentation and  
CC for diagnosis and treatment of prion diseases.  
XX  
SQ Sequence 685 AA;  
  
Query Match 100.0%; Score 701; DB 22; Length 685;  
Best Local Similarity 100.0%; Pred. No. 7.4e-64;  
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MSDSNOCNNQONTQOYSONGNOOQGNRRYQGYQAYNAQAPAGGYQNYQYSGYQCGY 60  
Db 1 msdngnqnqnyqysqngqgnmrlyqyqaynaqapaggyqnyqysqyqyqy 60  
  
QY 61 QQYPPDAGYQQOYYPGCGYQOQYFNPQGGYXOQFNPQGGYKFNFNNNNLOGYQAGFQ 120  
Db 61 qqynpddagyyqyppggyqyqyqyqyqyqyqyqyqyqyqyqyqyqyqyqy 120  
  
QY 121 SOG 123  
Db 121 sqg 123  
  
RESULT 2  
AAB30800 standard; Protein: 271 AA.  
XX  
AC AAB30800;  
XX  
DT 02-APR-2001 (first entry)  
XX  
DE A modified N region of yeast Sup35 protein.  
XX  
KM Sup35; phenotype: SCHAG; self-coalesces; higher-order aggregate;  
KM amyloidogenic domain; aggregation; fibril; phenotypic alteration;  
KM gene therapy; disease resistance; plant pigmentation; prion disease.  
XX  
OS Synthetic.  
OS Saccharomyces cerevisiae.  
XX  
PN WO200075324-A2.  
XX  
PD 14-DEC-2000.  
XX  
PF 09-JUN-2000; 2000MO-US15876.  
XX  
PR 09-JUN-1999; 99US-0138833.  
XX  
PA (ARCH-) ARCH DEV CORP.  
XX  
PI Lindquist S, Li L, Ma J, Liu J, Sondheimer N, Scheibel T;  
DR WPI; 2001-061723/07.  
DR N-PSDB; AAC86685.  
XX

PT New nucleic acid encoding chimeric proteins with self-assembly  
PT properties, useful e.g. for diagnosis and treatment of prion diseases,  
PT also related aggregates, fibrils and polymers -  
XX  
PS Claim 11; Page 135-136; 188bp; English.  
XX  
XX The present sequence represents a modified N region of Sup35 protein, in  
CC the second oligopeptide repeat has been expanded twice, creating a total  
CC of seven repeats. Sup35 possesses the prion-like capacity to undergo a  
CC self-perpetuating conformational alteration that changes the functional  
CC state of Sup35 in a manner that creates a heritable change in phenotype.  
CC It is used to construct chimeric polypeptides of the invention, which  
CC comprise at least one SCHAG (self-coalesces into higher-order aggregates)  
CC amino acid sequence fused in frame with a polypeptide of interest (which  
CC is other than a marker protein, a glutathione-S-transferase or a  
CC staphylococcal nuclear protein). The specification also describes  
CC chimeric polypeptides that comprises an amyloidogenic domain that causes  
CC aggregation into fibrils. The chimeric polypeptides are used to prepare  
CC polymers with multiple reactivities, e.g. derivatised with enzymes, or  
CC specific binding partners, and useful e.g. for performing multi-step  
CC chemical reactions. They can be used create an inducible, or stable  
CC phenotypic alteration in a cell, e.g. for gene therapy, protein  
CC production, imparting disease resistance to plants, altering plant  
CC pigmentation and for diagnosis and treatment of prion diseases.  
XX  
SQ Sequence 271 AA;  
  
Query Match 97.3%; Score 682; DB 22; Length 271;  
Best Local Similarity 87.2%; Pred. No. 2.1e-62;  
Matches 123; Conservative 0; Mismatches 0; Indels 18; Gaps 1;  
  
QY 1 MSDSNOCNNQONTQOYSONGNOOQGNRRYQGYQAYNAQAPAGGYQNYQYSGYQCGYQ--- 56  
Db 1 msdngnqnqnyqysqngqgnmrlyqyqaynaqapaggyqnyqysqyqyqy 60  
  
QY 57 -----QGGYQYNPDAQYQOQYNPQGGYQOQYFNPQGGYKFNFNNNNLOGYQAGFQ 102  
Db 61 qqynpddagyyqyppggyqyqyqyqyqyqyqyqyqyqyqyqyqyqyqyqy 120  
  
QY 103 NFNTNNNLOGYQAGFQPSOG 123  
Db 121 nfnynnmlggyqagfqpqsqg 141  
  
RESULT 3  
AAB30799 standard; Protein: 215 AA.  
XX  
ID AAB30799  
XX  
AC AAB30799;  
XX  
DT 02-APR-2001 (first entry)  
XX  
DE A modified N region of yeast Sup35 protein.  
XX  
KM Sup35; phenotype: SCHAG; self-coalesces; higher-order aggregate;  
KM amyloidogenic domain; aggregation; fibril; phenotypic alteration;  
KM gene therapy; disease resistance; plant pigmentation; prion disease.  
XX  
OS Synthetic.  
OS Saccharomyces cerevisiae.  
XX  
PN WO200075324-A2.  
XX  
PD 14-DEC-2000.  
XX  
PF 09-JUN-2000; 2000MO-US15876.  
XX  
PR 09-JUN-1999; 99US-0138833.  
XX  
PA (ARCH-) ARCH DEV CORP.  
XX  
PI Lindquist S, Li L, Ma J, Liu J, Sondheimer N, Scheibel T;  
XX







PR 14-MAY-1999; 990S-0134221.  
PR 14-MAY-1999; 990S-0134370.  
PR 18-MAY-1999; 990S-0134768.  
PR 19-MAY-1999; 990S-0134941.  
PR 20-MAY-1999; 990S-0135124.  
PR 21-MAY-1999; 990S-0135353.  
PR 24-MAY-1999; 990S-0135629.  
PR 25-MAY-1999; 990S-0136021.  
PR 27-MAY-1999; 990S-0136392.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0139119.  
PR 16-JUN-1999; 990S-0139452.  
PR 16-JUN-1999; 990S-0139453.  
PR 17-JUN-1999; 990S-0139492.  
PR 18-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139458.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.  
PR 21-JUN-1999; 990S-0139763.  
PR 21-JUN-1999; 990S-0139817.  
PR 22-JUN-1999; 990S-0139899.  
PR 23-JUN-1999; 990S-0140353.  
PR 23-JUN-1999; 990S-0140354.  
PR 24-JUN-1999; 990S-0140695.  
PR 28-JUN-1999; 990S-0140823.  
PR 29-JUN-1999; 990S-0140991.  
PR 30-JUN-1999; 990S-0141287.  
PR 01-JUL-1999; 990S-0141842.  
PR 01-JUL-1999; 990S-0142154.  
PR 02-JUL-1999; 990S-0142055.  
PR 06-JUL-1999; 990S-0142390.  
PR 08-JUL-1999; 990S-0142803.  
PR 09-JUL-1999; 990S-0142820.  
PR 12-JUL-1999; 990S-0142977.  
PR 13-JUL-1999; 990S-0143542.  
PR 14-JUL-1999; 990S-0143624.  
PR 15-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144085.  
PR 16-JUL-1999; 990S-0144086.  
PR 19-JUL-1999; 990S-0144325.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.  
PR 19-JUL-1999; 990S-0144333.  
PR 19-JUL-1999; 990S-0144334.  
PR 19-JUL-1999; 990S-0144335.  
PR 20-JUL-1999; 990S-0144352.  
PR 20-JUL-1999; 990S-0144632.  
PR 20-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0144814.  
PR 21-JUL-1999; 990S-0145086.  
PR 21-JUL-1999; 990S-0145088.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145089.  
PR 22-JUL-1999; 990S-0145192.  
PR 23-JUL-1999; 990S-0145145.  
PR 23-JUL-1999; 990S-0145218.  
PR 23-JUL-1999; 990S-0145224.  
PR 26-JUL-1999; 990S-0145276.  
  
PR 27-JUL-1999; 990S-0145913.  
PR 27-JUL-1999; 990S-0145918.  
PR 27-JUL-1999; 990S-0145919.  
PR 28-JUL-1999; 990S-0145951.  
PR 02-AUG-1999; 990S-0146386.  
PR 02-AUG-1999; 990S-0146388.  
PR 02-AUG-1999; 990S-0146389.  
PR 03-AUG-1999; 990S-0147038.  
PR 04-AUG-1999; 990S-0147204.  
PR 04-AUG-1999; 990S-0147302.  
PR 05-AUG-1999; 990S-0147192.  
PR 05-AUG-1999; 990S-0147260.  
PR 06-AUG-1999; 990S-0147303.  
PR 06-AUG-1999; 990S-0147416.  
PR 09-AUG-1999; 990S-0147493.  
PR 09-AUG-1999; 990S-0147935.  
PR 10-AUG-1999; 990S-0148171.  
PR 11-AUG-1999; 990S-0148319.  
PR 12-AUG-1999; 990S-0148341.  
PR 13-AUG-1999; 990S-0148565.  
PR 13-AUG-1999; 990S-0148684.  
PR 16-AUG-1999; 990S-0149368.  
PR 17-AUG-1999; 990S-0149175.  
PR 18-AUG-1999; 990S-0149426.  
PR 20-AUG-1999; 990S-0149722.  
PR 20-AUG-1999; 990S-0149723.  
PR 20-AUG-1999; 990S-0149929.  
PR 23-AUG-1999; 990S-0149902.  
PR 23-AUG-1999; 990S-0149930.  
PR 25-AUG-1999; 990S-0150566.  
PR 26-AUG-1999; 990S-0150884.  
PR 27-AUG-1999; 990S-0151065.  
PR 27-AUG-1999; 990S-0151066.  
PR 27-AUG-1999; 990S-0151080.  
PR 30-AUG-1999; 990S-0151303.  
PR 31-AUG-1999; 990S-0151438.  
PR 01-SEP-1999; 990S-0151930.  
PR 07-SEP-1999; 990S-0152363.  
PR 10-SEP-1999; 990S-0153070.  
PR 13-SEP-1999; 990S-0153758.  
PR 15-SEP-1999; 990S-0154018.  
PR 16-SEP-1999; 990S-0154039.  
PR 20-SEP-1999; 990S-0154779.  
PR 22-SEP-1999; 990S-0155139.  
PR 23-SEP-1999; 990S-0155486.  
PR 24-SEP-1999; 990S-0155486.  
PR 28-SEP-1999; 990S-0156458.  
PR 29-SEP-1999; 990S-0156596.  
PR 04-OCT-1999; 990S-0157117.  
PR 05-OCT-1999; 990S-0157753.  
PR 06-OCT-1999; 990S-0157865.  
PR 07-OCT-1999; 990S-0158029.  
PR 08-OCT-1999; 990S-0158232.  
PR 12-OCT-1999; 990S-0158369.  
PR 13-OCT-1999; 990S-0159293.  
PR 13-OCT-1999; 990S-0159294.  
PR 13-OCT-1999; 990S-0159295.  
PR 14-OCT-1999; 990S-0159329.  
PR 14-OCT-1999; 990S-0159330.  
PR 14-OCT-1999; 990S-0159331.  
PR 14-OCT-1999; 990S-0159637.  
PR 14-OCT-1999; 990S-0159637.  
PR 14-OCT-1999; 990S-0159638.  
PR 18-OCT-1999; 990S-0159584.  
PR 21-OCT-1999; 990S-0160741.  
PR 21-OCT-1999; 990S-0160767.  
PR 21-OCT-1999; 990S-0160768.  
PR 21-OCT-1999; 990S-0160770.  
PR 21-OCT-1999; 990S-0160814.  
PR 21-OCT-1999; 990S-0160815.  
PR 22-OCT-1999; 990S-0160980.  
PR 22-OCT-1999; 990S-0160981.  
PR 22-OCT-1999; 990S-0160989.  
PR 25-OCT-1999; 990S-0161404.



PR	14-MAY-1999	9905-0134219
PR	14-MAY-1999	9905-0134221
PR	14-MAY-1999	9905-0134370
PR	18-MAY-1999	9905-0134768
PR	19-MAY-1999	9905-0134841
PR	20-MAY-1999	9905-0135124
PR	21-MAY-1999	9905-0135523
PR	24-MAY-1999	9905-0135629
PR	25-MAY-1999	9905-0136021
PR	27-MAY-1999	9905-0136392
PR	28-MAY-1999	9905-0136782
PR	01-JUN-1999	9905-0137222
PR	03-JUN-1999	9905-0137528
PR	04-JUN-1999	9905-0137502
PR	07-JUN-1999	9905-0137724
PR	08-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138540
PR	10-JUN-1999	9905-0138847
PR	14-JUN-1999	9905-0139119
PR	16-JUN-1999	9905-0139452
PR	16-JUN-1999	9905-0139493
PR	17-JUN-1999	9905-0139492
PR	18-JUN-1999	9905-0139454
PR	18-JUN-1999	9905-0139455
PR	18-JUN-1999	9905-0139456
PR	18-JUN-1999	9905-0139457
PR	18-JUN-1999	9905-0139458
PR	18-JUN-1999	9905-0139459
PR	18-JUN-1999	9905-0139460
PR	18-JUN-1999	9905-0139461
PR	18-JUN-1999	9905-0139462
PR	18-JUN-1999	9905-0139463
PR	23-JUN-1999	9905-0140354
PR	24-JUN-1999	9905-0140695
PR	28-JUN-1999	9905-0140823
PR	29-JUN-1999	9905-0140931
PR	30-JUN-1999	9905-0141287
PR	01-JUL-1999	9905-0141244
PR	01-JUL-1999	9905-0141852
PR	02-JUL-1999	9905-0142050
PR	06-JUL-1999	9905-0142290
PR	08-JUL-1999	9905-0142803
PR	09-JUL-1999	9905-0142927
PR	12-JUL-1999	9905-0143297
PR	13-JUL-1999	9905-0143524
PR	14-JUL-1999	9905-0143542
PR	15-JUL-1999	9905-0144005
PR	16-JUL-1999	9905-0144085
PR	16-JUL-1999	9905-0144086
PR	19-JUL-1999	9905-0144325
PR	19-JUL-1999	9905-0144331
PR	19-JUL-1999	9905-0144332
PR	19-JUL-1999	9905-0144333
PR	19-JUL-1999	9905-0144334
PR	19-JUL-1999	9905-0144335
PR	20-JUL-1999	9905-0144352
PR	20-JUL-1999	9905-0144632
PR	20-JUL-1999	9905-0144684
PR	21-JUL-1999	9905-0144814
PR	21-JUL-1999	9905-0145088
PR	22-JUL-1999	9905-0145088
PR	22-JUL-1999	9905-0145087
PR	22-JUL-1999	9905-0145089
PR	22-JUL-1999	9905-0145192
PR	23-JUL-1999	9905-0145145
PR	23-JUL-1999	9905-0145218
PR	23-JUL-1999	9905-0145224

PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145921.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 22-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160880.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 21.1%; Score 148; DB 21; Length 419;  
Best Local Similarity 34.8%; Pred. No. 2.8e-07;  
Matches 57; Conservative 9; Mismatches 42; Indels 56; Gaps 13;  
  
QY 11 ONYOOYSONGNO---OQCNNRY---QCY---QAYMAQAP--AGGYON-----YQG 51  
Db 215 gnfgrntqyqgppmqgsgyspqqyatlppqgqgtatppfgygynqgprpppyq- 273  
QY 52 YSGYQOG-----GYOQYN-----PDAGYOOOYVPOGXYOQYN--POGXYOQ----- 90  
Db 274 -agynqgqgspvpppyqgynvqgspvpppygltqssyq-qqsgsnysagqpqgynqgqpr 331  
QY 91 QFNPOGGKGNKYNFNNNL-----QGYQAGFQPSQ 122  
Db 332 nypnpgq-gnfgpasgagnlgpapqagupgyqgyspqpqgqnd 374  
  
RESULT 11  
AAG30011  
ID AAG30011 standard; Protein: 439 AA.  
XX  
AC AAG30011;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 35804.  
XX  
KW Protein identification: signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 9905-0134218.  
PR 14-MAY-1999; 9905-0134219.  
PR 14-MAY-1999; 9905-0134221.  
PR 14-MAY-1999; 9905-0134370.  
PR 18-MAY-1999; 9905-01343768.  
PR 19-MAY-1999; 9905-0134941.  
PR 20-MAY-1999; 9905-0135124.  
PR 21-MAY-1999; 9905-0135353.  
PR 24-MAY-1999; 9905-0135629.  
PR 25-MAY-1999; 9905-0136021.  
PR 27-MAY-1999; 9905-0136392.  
PR 28-MAY-1999; 9905-0136782.  
PR 01-JUN-1999; 9905-0137222.  
PR 03-JUN-1999; 9905-0137528.  
PR 04-JUN-1999; 9905-0137502.  
PR 07-JUN-1999; 9905-0137724.  
PR 08-JUN-1999; 9905-0138094.  
PR 10-JUN-1999; 9905-0138540.  
PR 10-JUN-1999; 9905-0138847.  
PR 14-JUN-1999; 9905-0139119.  
PR 16-JUN-1999; 9905-0139452.  
PR 16-JUN-1999; 9905-0139453.  
PR 17-JUN-1999; 9905-0139492.  
PR 18-JUN-1999; 9905-0139454.  
PR 18-JUN-1999; 9905-0139455.  
PR 18-JUN-1999; 9905-0139456.  
PR 18-JUN-1999; 9905-0139457.  
PR 18-JUN-1999; 9905-0139458.  
PR 18-JUN-1999; 9905-0139459.  
PR 18-JUN-1999; 9905-0139460.  
PR 18-JUN-1999; 9905-0139461.  
PR 18-JUN-1999; 9905-0139462.  
PR 18-JUN-1999; 9905-0139463.  
PR 18-JUN-1999; 9905-0139750.  
PR 18-JUN-1999; 9905-0139763.  
PR 21-JUN-1999; 9905-0139817.  
PR 22-JUN-1999; 9905-0139819.  
PR 23-JUN-1999; 9905-0140353.  
PR 23-JUN-1999; 9905-0140354.  
PR 24-JUN-1999; 9905-0140695.  
PR 28-JUN-1999; 9905-0140823.  
PR 29-JUN-1999; 9905-0140991.  
PR 30-JUN-1999; 9905-0141287.  
PR 01-JUL-1999; 9905-0141842.  
PR 01-JUL-1999; 9905-0142154.  
PR 02-JUL-1999; 9905-0142055.  
PR 06-JUL-1999; 9905-0142390.  
PR 08-JUL-1999; 9905-0142803.  
PR 09-JUL-1999; 9905-0142920.  
PR 12-JUL-1999; 9905-0142977.  
PR 13-JUL-1999; 9905-0143542.  
PR 14-JUL-1999; 9905-0143624.  
PR 15-JUL-1999; 9905-0144005.  
PR 16-JUL-1999; 9905-0144085.  
PR 16-JUL-1999; 9905-0144086.  
PR 19-JUL-1999; 9905-0144325.  
PR 19-JUL-1999; 9905-0144331.  
PR 19-JUL-1999; 9905-0144332.  
PR 19-JUL-1999; 9905-0144333.  
PR 19-JUL-1999; 9905-0144334.  
PR 19-JUL-1999; 9905-0144335.  
PR 20-JUL-1999; 9905-0144352.  
PR 20-JUL-1999; 9905-0144632.  
PR 20-JUL-1999; 9905-0144884.  
PR 21-JUL-1999; 9905-0144814.  
PR 21-JUL-1999; 9905-0145086.  
PR 21-JUL-1999; 9905-0145088.  
PR 22-JUL-1999; 9905-0145085.  
PR 22-JUL-1999; 9905-0145087.  
PR 22-JUL-1999; 9905-0145089.  
PR 22-JUL-1999; 9905-0145192.  
PR 23-JUL-1999; 9905-0145145.  
PR 23-JUL-1999; 9905-0145218.

PR 23-JUL-1999; 9905-0145224.  
PR 26-JUL-1999; 9905-0145276.  
PR 27-JUL-1999; 9905-01452913.  
PR 27-JUL-1999; 9905-0145918.  
PR 27-JUL-1999; 9905-0145919.  
PR 28-JUL-1999; 9905-0145951.  
PR 02-AUG-1999; 9905-0146386.  
PR 02-AUG-1999; 9905-0146388.  
PR 03-AUG-1999; 9905-0147038.  
PR 04-AUG-1999; 9905-0147204.  
PR 04-AUG-1999; 9905-0147302.  
PR 05-AUG-1999; 9905-0147192.  
PR 06-AUG-1999; 9905-0147260.  
PR 06-AUG-1999; 9905-0147303.  
PR 06-AUG-1999; 9905-0147416.  
PR 09-AUG-1999; 9905-0147493.  
PR 09-AUG-1999; 9905-0147935.  
PR 10-AUG-1999; 9905-0148171.  
PR 11-AUG-1999; 9905-0148319.  
PR 12-AUG-1999; 9905-0148341.  
PR 13-AUG-1999; 9905-0148565.  
PR 13-AUG-1999; 9905-0148684.  
PR 16-AUG-1999; 9905-0149368.  
PR 17-AUG-1999; 9905-0149175.  
PR 18-AUG-1999; 9905-0149426.  
PR 20-AUG-1999; 9905-0149722.  
PR 20-AUG-1999; 9905-0149723.  
PR 20-AUG-1999; 9905-0149929.  
PR 23-AUG-1999; 9905-0149902.  
PR 23-AUG-1999; 9905-0149930.  
PR 25-AUG-1999; 9905-0150566.  
PR 26-AUG-1999; 9905-0150884.  
PR 27-AUG-1999; 9905-0151065.  
PR 27-AUG-1999; 9905-0151066.  
PR 27-AUG-1999; 9905-0151080.  
PR 30-AUG-1999; 9905-0151303.  
PR 31-AUG-1999; 9905-0151438.  
PR 01-SEP-1999; 9905-0151930.  
PR 07-SEP-1999; 9905-0152363.  
PR 10-SEP-1999; 9905-0153070.  
PR 13-SEP-1999; 9905-0153758.  
PR 15-SEP-1999; 9905-0154018.  
PR 16-SEP-1999; 9905-0154039.  
PR 20-SEP-1999; 9905-0154779.  
PR 22-SEP-1999; 9905-0155139.  
PR 23-SEP-1999; 9905-0155486.  
PR 24-SEP-1999; 9905-0155659.  
PR 28-SEP-1999; 9905-0156458.  
PR 29-SEP-1999; 9905-0156596.  
PR 04-OCT-1999; 9905-0157117.  
PR 05-OCT-1999; 9905-0157753.  
PR 06-OCT-1999; 9905-0157865.  
PR 07-OCT-1999; 9905-0158029.  
PR 08-OCT-1999; 9905-0158232.  
PR 12-OCT-1999; 9905-0158369.  
PR 13-OCT-1999; 9905-0159293.  
PR 13-OCT-1999; 9905-0159294.  
PR 13-OCT-1999; 9905-0159295.  
PR 14-OCT-1999; 9905-0159329.  
PR 14-OCT-1999; 9905-0159330.  
PR 14-OCT-1999; 9905-0159331.  
PR 14-OCT-1999; 9905-0159637.  
PR 14-OCT-1999; 9905-0159638.  
PR 18-OCT-1999; 9905-0159584.  
PR 21-OCT-1999; 9905-0160741.  
PR 21-OCT-1999; 9905-0160767.  
PR 21-OCT-1999; 9905-0160768.  
PR 21-OCT-1999; 9905-0160770.  
PR 21-OCT-1999; 9905-0160814.  
PR 21-OCT-1999; 9905-0160815.  
PR 22-OCT-1999; 9905-0160980.  
PR 22-OCT-1999; 9905-0160981.





XX Tapetum-specific promoter encoded by clone pMS10.  
 DE Maize; sterility.  
 XX  
 KM WO9008831-A.  
 XX  
 PN 09-AUG-1990.  
 XX  
 PD 26-JAN-1990; 90MO-GH00114.  
 XX  
 PF 26-JAN-1989; 89GB-0001675.  
 XX  
 PR (ICIL ) IMPERIAL CHEM INDS PLC.  
 XX  
 PA Bridges IG, Bright SMJ, Greenland AJ, Schuch WW, Reid GA;  
 PI WPI, 1990-260940/34.  
 DR N-PSDB; AAQ05759.  
 DR  
 XX  
 PT Inhibition of gene expression in plants - which is useful in e.g.  
 PT imparting male sterility in maize.  
 PS Disclosure; Fig 1; 40pp; English.  
 XX  
 CC DNA encoding the promoter is used in a gene construct with a dis-  
 CC rupter gene such as the mammalian uncoupling protein gene, or a  
 CC mutated form of the gene for the beta subunit of F1-ATPase. The  
 CC promoter operates in the target tissue and expression of the dis-  
 CC rupter gene inhibits respiration resulting in cell death. This is  
 CC of particular use for inhibiting male fertility in maize crops  
 CC where cell death results in failure to produce viable pollen. When  
 CC inserted into a gene cascade, as a module which permits external  
 CC control of expression, fertility may also be restored if required.  
 CC See also AAR06524 and AAR06525.  
 XX  
 SQ Sequence 341 AA;

Query Match 19.1%; Score 134; DB 11; Length 341;  
 Best Local Similarity 30.9%; Pred. No. 5.9e-06;  
 Matches 43; Conservative 11; Mismatches 55; Indels 30; Gaps 7;  
 QY 2 SDSNQGNNQONQYQYSONGNOQOGNNRNYQGYAOAYNAOAPAGGYQ-----NYQGYSGY 55  
 Db 126 srpcttrntrnyrgnygdgppqgn-----yq--ntrppegyqgnppqgnyqlyrsq 178  
 QY 56 QQGGYQOQYNDAGYQ-----QYNPQGGYQ--QYNPQGGYQOQFNPPQGGRCGY 101  
 Db 179 qdg--rgyapqnyagggqdgqrgfgrndyldrsyngpdtfrsqtqyqgnvnpagqgqgy 236  
 QY 102 KNFNYNNNNL-QGYOAGFQP 119  
 Db 237 nnpqertlfsqgqgqgfrp 255

Search completed: February 1, 2002, 15:01:49  
 Job time: 254 sec



RESULT 2  
ID 020497 PRELIMINARY: PRT: 3498 AA.  
AC 020497;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE F47M4.2 PROTEIN.  
GN F47M4.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Shownkeen R.,  
RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierri-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Welnsstock L., Wilkinson-Spiro J., Wohlman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
DR EMBL: 249888; CAAG0064.1; -  
DR InterPro: IPR000504; RRM-like.  
DR InterPro: IPR000561; EGF-like.  
DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
DR PROSITE: PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
SQ SEQUENCE 3498 AA; 400778 MW; 283F6C5A6B9E336 CRC64;

Query Match 17.7%; Score 239; DB 5; Length 3498;  
Best Local Similarity 31.4%; Pred. No. 1e-12;  
Matches 101; Conservative 20; Mismatches 117; Indels 84; Gaps 17;  
OY 1 OGGGG--GGGGGGGGGG--GGGFTALASSFMNSNNNN-- 43  
DB 3033 OGTGTGAGGAGGCTGGGGGGGGGTP--AQQQPPNOYSSNOQVQQAQQQQQPLNQ 3091  
OY 44 -NOSGGSFSGALASSFMNSNNNNNSNNSQGYNSYONGNONSQGYNNQOYOGGNG 102  
DB 3092 VSGSGAAGGGR-PGSDAAYQSGYNGTG-----NQSRYRPDQGGGQAGNQMGSSNQ 3143  
OY 103 GYVQ-----QGSGGAFSSLAS-----MAQS--YLG-GGOTOSNOQOYMOGO 143  
DB 3144 AONQLRSQAQAQPLQPPQSOQFPQPAQAQAKNPMASQAYGCGFGGQQQGYDQQ--QGGQ 3201  
OY 144 NNQOQYQ--QGQNYQHQQGQQGQGHSSSFALASSMSS-----YLGK 186  
DB 3202 IAPQAQANPQASQSY-GQQQTQONRYGKSSGTTANSGGSSNILLNOSMEESGILNQGFSGA 3260  
OY 187 NSNSNSYSYG--QQANNEYGRPHNQOOSNEYGRPOYG-----GNONSNGOHS--- 234  
DB 3261 SSVASSQGGSSGQMGSGYGMGPMQMGQMGQMGQMGQMGQMGQMGQMGQMGQMGQMG 3320  
OY 235 -----FNFSGNFSQNNNGNQ 250  
DB 3321 QTGMSRSGLGGIGGGGQSOQ 3342

RESULT 3  
ID 019662 PRELIMINARY: PRT: 382 AA.

AC 019662;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE SIMILARITY TO E. COLI CELL DIVISION PROTEIN FTSK.  
GN F2IC10.8.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Shownkeen R.,  
RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierri-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Welnsstock L., Wilkinson-Spiro J., Wohlman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX Du Z., Gattung S.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;  
RA Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U55364; AAA97972.1; -  
KW Cell division.  
SQ SEQUENCE 382 AA; 42877 MW; 3BD9EB544BBFB2C CRC64;

Query Match 17.6%; Score 237; DB 5; Length 382;  
Best Local Similarity 28.7%; Pred. No. 2e-13;  
Matches 83; Conservative 32; Mismatches 92; Indels 82; Gaps 13;  
OY 15 OGGGGFTALASSFMNSNNNNQGGGSSFGALASSFMNSNNNNNS 74  
DB 73 QNQQYGS-----TAQYNNQNMQNMQNMQNMQNMQNMQNMQNMQNMQNMQNMQNM 112  
OY 75 QGGYNGSY-----QNGNONS-----QGYNNQOYGGNGGYQQQSGGAFSSLASMA 122  
DB 113 QGCGQNNYGTQAOYDAGQNMQMGSGGYGYSNNGYSNMQNSQMSQSSSG--SSSVSLM 169  
OY 123 QSYLGGGTQNSNOQYNNQGNQ--QYQOQGGQNTQH--QGGQOQQQGH-----SSSFS 174  
DB 170 DYSFNNNGNCQYKDGQYIENGQTRQATQOELQIQYRNSVDQYMNQYGVGQWINSVFQ 229  
OY 175 ALASSMSSY-----LGNNNSNS-----SYGQ-----QQ 199  
DB 230 TLPSYNNSPFNIPMTPTAPCLCNAQNGNTTNTNATGINTPQARDAYGNQMSNSOFOQ 289  
OY 200 ANEYGRPHNG--QQSNEYGRPHNGNONSNGOHSFNFSGNFSQNNN 247  
DB 290 ANONGQYNTQYQOQNGQNGQNF-----QNSNGQNMNSQYQQTSTNOQNMN 335

RESULT 4  
ID 019414 PRELIMINARY: PRT: 409 AA.  
AC 019414;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)









AC 044011;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE PROTEIN KINASE YAKA.  
 GN YAKA.  
 OS Dictyostellium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.  
 NCBI\_Taxid=44689;  
 RN (1)  
 RP STRAIN=AK800;  
 RC MEDLINE=96042901; PubMed=8536963;  
 RX Loomis W.F., Welker D., Hughes J., Maghakian D., Kuspa A.;  
 RA "Integrated maps of the chromosomes in Dictyostellium discoideum."; Genetics 141:147-157(1995).  
 RL (12)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AK800;  
 RX MEDLINE=96224325; PubMed=8643615;  
 RA Kuspa A., Loomis W.F.;  
 RT "Ordered yeast artificial chromosome clones representing the Dictyostellium discoideum genome."; Proc. Natl. Acad. Sci. U.S.A. 93:5562-5566(1996).  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AK800;  
 RA Kuspa A., Lu S., Souza G.M.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AF045453; AAC02554.1; -  
 DR HSSP; P24941; 1CKP.  
 DR InterPro: IPR000719; Euk.pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_kin\_actsite.  
 DR Pfam: PF00069; pkinase; 2.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS0108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; Serine/threonine-protein kinase; Transferase.  
 SO SEQUENCE 1457 AA; 167110 MW; C1FCDC99D561856 CRC64;

Query Match 16.2%; Score 218.5; DB 5; Length 1457;  
 Best Local Similarity 29.3%; Pred. No. 2.9e-11;  
 Matches 77; Conservative 25; Mismatches 98; Indels 63; Gaps 9;  
 QY 16 GCGGGSFTALASLASS-----FMNSNNNOGQNGSSGSGFALASMASSFM 63  
 DB 789 GSSWGDSSISLNPSTPTQKMFQOOQYNNNNNNNNNNNNNNNNNNNNNN-- 846  
 QY 64 HSNNNONSNS-----QCYNSYNGNGNSGYNNOQYQCG 100  
 DB 847 NNNNNNNRRNRKSDIPDSFSSSEGMDFNLVYQOOQOOQOOQOOQOOQOO 906  
 QY 101 NGCYQOQGGSGGAFSSLASM-----AGSYLGGGCTOSNQOYNOOGON 144  
 DB 907 QO 964  
 QY 145 NQOQYQO 203  
 DB 965 NQO 1023  
 QY 204 GRPO---HNGGQOSNEY---GRP 220  
 DB 1024 KQOQKMLHHQO 1046

RESULT 12  
 ID 041553 PRELIMINARY; PRT; 815 AA.  
 AC 041553;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE TRITICUM AESTIVUM HIGH MW GLUTENIN SUBUNIT (AX2\*).  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
 OC Triticeae; Triticum.  
 NCBI\_Taxid=4565;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. CHEYENNE;  
 RA Anderson O.D., Greene F.C.;  
 RL Theor. Appl. Genet. 77:689-700(1989).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. CHEYENNE;  
 RA Anderson O.D.;  
 RL Submitted (JAN-1989) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. CHEYENNE;  
 RA Anderson O.D.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; M22208; AAB02788.1; -  
 DR Mendel; 14743; Trlae; 2380; 14743.  
 DR InterPro: IPR001419; Glutennin.  
 DR PRINTS; PRO0210; GLUTENIN.  
 SO SEQUENCE 815 AA; 88476 MW; FC09905398EA71A8 CRC64;

Query Match 16.2%; Score 218; DB 10; Length 815;  
 Best Local Similarity 31.9%; Pred. No. 1.9e-11;  
 Matches 87; Conservative 13; Mismatches 125; Indels 48; Gaps 12;  
 QY 2 GCGGCGGCGG-0GCGGCGGGSFTALASLASSFMNSNNNOGQNGSSGSGFALASMAS 60  
 DB 198 GCGGQLRGCGGQO 248  
 QY 61 SFMNSNNONSNSQOQY---NQSYNGNGNSGYN-----NQYQGGNG--GY----- 104  
 DB 249 --QOQSGGQO 305  
 QY 105 -QOQOQGGGAFSSLASMAQSYLGGGCTOSNQOYNOQ-GONNOQYQOQOQOQOQO 162  
 DB 306 QOQSGGSGGYPRPSAQPGQLQSTQEQQLQEQOQOQOQOQOQOQOQOQOQOQO 363  
 QY 163 QOQOQGHSSSFALASMASSYLGNNSNSSSYGGQOQOQOQOQOQOQOQOQOQO 221  
 DB 364 QGQOPGQRP-----GYSTSPQQLGCGGPRYPTSPQPGQEQOQOQOQOQO 411  
 QY 222 YG---GNQNSNGHSEFNGFNFSQONNGNQR 252  
 DB 412 QGQOQOQPEQGQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 444

RESULT 13  
 ID 003872 PRELIMINARY; PRT; 830 AA.  
 AC 003872;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE HIGH MOLECULAR WEIGHT GLUTENIN SUBUNIT 1A1 PRECURSOR.  
 GN GLU-1A1.  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
 OC Triticeae; Triticum.  
 NCBI\_Taxid=4565;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. HOPE; TISSUE=ENDOSPERM;  
 RA Halford N.G., Field J.M., Blair H., Urwin P., Moore K., Robert L.,















```
CC -1- SUBUNIT: COMPONENT OF NUCLEAR RIBOPROTEIN COMPLEXES.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG ISOFORM (SHOWN HERE)
CC AND A SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBQUITOUS.
CC -1- DISEASE: A FORM OF MALIGNANT MYXOID LIPOSARCOMA IS CHARACTERIZED
CC BY A CHROMOSOMAL TRANSLOCATION T(12;16)(Q13;P11) THAT INVOLVES FUS
CC AND CHOP.
CC -1- DISEASE: A FORM OF ACUTE MYELOID LEUKEMIA (AML) IS CHARACTERIZED
CC BY A CHROMOSOMAL TRANSLOCATION T(16;21)(P11;Q22) THAT INVOLVES
CC FUS AND EGR.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 RANBP-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: S62140; AB27102.1; ALT_SEQ.
DR EMBL: S62138; AB27103.1; ALT_SEQ.
DR EMBL: X71427; CA50558.1; ALT_SEQ.
DR EMBL: X71428; CA50559.1; ALT_SEQ.
DR EMBL: AF071213; AAC35285.1; -.
DR EMBL: AF071213; AAC35284.1; -.
DR PIR: S33799; S33799.
DR HSSP: P09651; 1HA1.
DR MIM: 137070; -.
DR InterPro: IPR000504; RRM.
DR InterPro: IPR001876; Znf-RanBP.
DR Pfam: PF00076; rrm; 1.
DR Pfam: PF00641; zt-RanBP; 1.
DR SMART: SM00360; RRM; 1.
DR SMART: SM00547; Znf_RB2; 1.
DR PROSITE: PS50102; RRM; 1.
DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
KW Proto-oncogene; RNA-binding; DNA-binding; Nuclear protein; Repeat;
KW Alternative splicing; Chromosomal translocation; Zinc-finger; Zinc;
KW Metal-binding.
FT DOMAIN 1 165 GLN/GLY/SER/TYR-RICH.
FT DOMAIN 166 267 GLY-RICH.
FT DOMAIN 285 371 RNA-BINDING (RRM).
FT DOMAIN 371 526 ARG/GLY-RICH.
FT ZN_FING 422 453 RANBP-TYPE.
FT SITE 266 267 BREAKPOINT FOR TRANSLOCATION TO FORM
FT VARSPLIC 64 65 FUS/TLS-CHOP ONCOGENE.
FT CONFLICT 338 338 TG -> S (IN SHORT ISOFORM).
FT FT SEQUENCE 526 AA; 53426 MW; 88C8E263B7905549 CRC64;
```

Query Match 16.0%; Score 216; DB 1; Length 526;  
Best Local Similarity 30.5%; Pred. No. 2, 7e-07;  
Matches 65; Conservative 15; Mismatches 79; Indels 54; Gaps 6;

```
OY 33 MNSNNNNNGGONSGGSGFGLASSMASSFMHSNNNNNSNGQYNGSYQNGNNSGQY 92
DB 1 MASNDYTOAT-----QSAYPTOPG-----QGSYQ-----QSSQPY 33
OY 93 NNNOYGGNGGYYQQQGGGAFSSILASMAQSYLGGGOTQSNNOQYNNQSQNNQYQQQ 152
DB 34 GQOSTY---SGTSQSTDTSTGYGQSYSTYGQSONTGITQSTPQGGTGGYSSQSSQS 89
OY 153 GONYGHOQGGGQGGGSSFSALASMASSYLLNNSNSNSSYGQGOQANEGRPHNGQQ 212
DB 90 SYGGQSSSYGYGGQPPAPSS-----TSSYSSSSSSSSSYGGQPOSSSY 131
OY 213 QSNETGRPOYGGNNSNGCHESFNSGNSFNSQON 245
```

DB 132 Q-----QPSYGGGQSSYGGQSSYNPPOGYGOQN 159

```
RESULT 7
ID N100_YEAST STANDARD; PRT; 959 AA.
AC 002629;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE NUCLEOPORIN NUP100/NSP100 (NUCLEAR PORE PROTEIN NUP100/NSP100).
GN NUP100 OR NSP100 OR YKL068W OR YKLJ36.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054906; PubMed=1385442;
RA Wente S.R., Rout M.P., Blobel G.;
RT "A new family of yeast nuclear pore complex proteins.";
RL J. Cell Biol. 119:705-723(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94378724; PubMed=8091863;
RA Rasmussen S.W.;
RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
RT NUP100 gene, an open reading frame (ORF) possibly representing a
RT nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
RT addition to seven ORFs with weak or no significant similarity to
RT known proteins.";
RL Yeast 10:569-574(1994).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLOCATION OF
CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX.
CC -1- DOMAIN: CONTAINS G-L-F-G REPEATS.
CC -1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: Z15035; CAAT8753.1; -.
DR EMBL: X75780; CAAS3406.1; -.
DR EMBL: Z28068; CAAB1905.1; -.
DR PIR: B44402; B44402.
DR PIR: S39173; S39173.
DR PIR: S44518; S44518.
DR SGD: S000151; NUP100.
KW Nuclear protein; Transport; Repeat.
FT DOMAIN 33 571 G-L-F-G.
FT FT SEQUENCE 959 AA; 99988 MW; D3985F9901BBA51 CRC64;
```

Query Match 16.0%; Score 215.5; DB 1; Length 959;  
Best Local Similarity 28.0%; Pred. No. 5, 1e-07;  
Matches 84; Conservative 37; Mismatches 110; Indels 69; Gaps 15;

```
OY 2 GGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 43
DB 36 GNNNNNNNNSTNNAQSGGCGGFTSAAGSNSNSLFGNNNTQNNGAFGGSGMATONSPPGSLN 95
OY 44 -NNSGSSFGALASMASSFMHS-----NNNNNSNSGQGYNOSYONG-----NNSGQY 92
DB 96 SSNNSNNGNFTGCGSSSMGSGFCGNTNNAFNNNSNSTNSPFGNKPVTGGTLFGSONNNNSAG- 154
```







[illegible][illegible]



0y 207 OHNGGQSNENYGRPOLYGNNONSNGHESNFNGSNNNNGNO 250  
Db 479 ---GGRRPQGCGQHPEGCGQPEGSGGQGYTPPS---PDPRGSGOO 516

RESULT 15

RB56\_HUMAN STANDARD: PRT; 592 AA.

ID RB56\_HUMAN  
AC 092804; 092751;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DE 20-AUG-2001 (Rel. 40, Last annotation update)  
DE TATA-BINDING PROTEIN ASSOCIATED FACTOR 2N (RNA-BINDING PROTEIN 56) (TAFII168) (TAF(II)68).  
GN TAF2N OR RBP56.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).  
RX MEDLINE=97124846; PubMed=8954779.  
RA Morohoshi F., Arai K., Takahashi E.-I., Tanigami A., Ohki M.;  
RT "Cloning and mapping of a human RBP56 gene encoding a putative RNA binding protein similar to FUS/TLS and EMS proteins.";  
RL Genomics 38:51-57(1996).  
[2]  
SEQUENCE FROM N.A. (SHORT ISOFORM), AND SEQUENCE OF 282-297 & 307-320.  
RX MEDLINE=97045110; PubMed=8890175;  
RA Bertoletti A., Lutz Y., Heard D.J., Chambon P., Tora L.;  
RT "hRAF(II)68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EMS is associated with both TFIIID and RNA polymerase II.";  
RL EMBO J. 15:5022-5031(1996).  
[3]  
SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).  
RX MEDLINE=99013873; PubMed=9795213;  
RA Morohoshi F., Ootsuka Y., Arai K., Ichikawa H., Mitani S., Munakata M., Ohki M.;  
RT "Genomic structure of the human RBP56/hRAFII68 and FUS/TLS genes.";  
RL Gene 221:191-198(1998)  
CC -1- FUNCTION: RNA AND ssDNA-BINDING PROTEIN THAT MAY PLAY SPECIFIC ROLES DURING TRANSCRIPTION INITIATION AT DISTINCT PROMOTERS. CAN ENTER THE PREINITIATION COMPLEX TOGETHER WITH THE RNA POLYMERASE II (POL II).  
CC -1- SUBUNIT: BELONGS TO THE RNA POLYMERASE II (POL II) TRANSCRIPTIOANAL MULTIPROTEIN COMPLEX, TOGETHER WITH THE TATA-BINDING PROTEIN (TBP) AND OTHER TBP-ASSOCIATED FACTORS (TAF(II)S).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG ISOFORM (SHOWN HERE) AND A SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: UBIQUITOUS. OBSERVED IN ALL FETAL AND ADULT TISSUES.  
CC -1- DISEASE: A FORM OF EXTRASKELETAL MYXOID CHONDROSARCOMAS (EMC) IS CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(9;17)(Q22;Q11) THAT INVOLVES TAF2N AND NR4A3.  
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNM).  
CC -1- SIMILARITY: CONTAINS 1 RAMBP-TYPE ZINC FINGER.  
CC -1- SIMILARITY: BELONGS TO THE TEF FAMILY OF RNP PROTEINS.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U51334; AAC50932.1; -  
DR EMBL: X98893; CAA67398.1; -  
DR EMBL: AB010067; BAA33811.1; -  
DR EMBL: AB010057; BAA33811.1; JOINED

[illegible]



Db 61 SGYSOSYGGYENOKOSYSQPPYNNQGGQQNNMESSGGGGRAPSYDQDPY-GQQDSYDQQ 119  
OY 108 QGOSGAFSSILASMAOSYLGCGOTOQSNQOQYNNQGGNNQOQYQOQGGANYQHQQGGQQQQQ 167  
Db 120 SGY-----DQHGSYDEQSNYDQHDYSQNOQSYHSQRENTSHHTQDDRD- 166  
OY 168 GHSSFSALASMASSYLGNNNSNSSSYGGQQQANEX--GRPOHNGQQQSNEXGRPOYGG 224  
Db 167 -----VSRYGEDNRGSGGQGGGRCRGYDKDGRGPMTGSSGGDRGCFKFGG 214  
OY 225 NQN-----SNGHESFNFSGN 240  
Db 215 HRDYGPRTDASESDNSDNN 234

Search completed: February 1, 2002, 15:07:33  
Job time: 313 sec





Db	333	SSYLGNNSNSMSSYGGGQQAANEYRPFQHNQGSNEFYGRPYGNGNSNGQHESEFNFSGN	392
Qy	241	FSQOINNGNQRY	253
Db	393	FSQOINNGNQRY	405

RESULT 2  
T22330  
hypothetical protein F47M4.2 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T22330  
R:Morrison, B.  
submitted to the EMBL Data Library, June 1995

A:Cross-references: EMBL:Z49888; PIDD:CA90064.1; GSPDB:GN00028; CESP:F47A4.2  
A:Experimental source: clone F47A4  
C:Genetics:  
A:Gene: CESP:F47A4.2  
A:Map position: X  
A:Introns: 36/3; 72/3; 147/2; 228/3; 261/2; 320/2; 847/3; 881/1; 1317/2; 1500/3; 1653/2.

[illegible]

```

RESULT      3
T29339
hypothetical protein F21C10.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29339
R:Du, Z.; Galtung, S.
submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid F21C10.
A:Reference number: Z20610
A:Accession: T29339
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1*382 <DUZ>
A:Cross-references: EMBL:U55364; PIDN:AA97972.1; GSPDB:GN00023; CESP:F21C10.8

```

A: Experimental source: strain Bristol N2; clone F21C10  
C: Genetics:  
A: Gene: CESP: F21C10.8  
A: Map position: 5  
A: Introns: 20/1; 219/3; 249/3

Query Match	17.6%;	Score 237;	DB 2;	Length 382;
Best Local Similarity	28.7%;	Pred. No. 3e-09;		
Matches	83;	Mismatches	92;	Indels 82;
				Gaps 13;

```

QY 15 0G0GGGFTTALASLASFSFMSNNNNNOG0N0SSGSSSPGALASMASSFFHSHNNNNNSNS 74
Db 73 0NQ0YGS-----TAQNNQNM0NM0NSNSNYSTTNN-----NNQ0NSNQY 112
QY 75 00GY0MST-----0NG0N0NS-----0GYNNQY0GGNGCY000G0SGAFSLSMA 122
Db 113 00G000NNNGT0A0YDAG0NM0G0SG0Y0SNM0YSNM0NS0MS0QAS0SG-----SSYS1LM 169
QY 123 0SYLGGG0T0S000Y0N0G0G0NN0--00Y000G0GY0HQ--00G0Q0Q00GH-----SSFS 174
Db 170 DYSPNNGM0QTKD0GYEN0GT0R0AT00EL0Q100YKRISYDQYNN0VNGYGM0INSVF0 229
QY 175 ALASMASSY-----LGNNSNSNS-----SYGQ0-----QQ 199
Db 230 TLPSVNSNFSN1PTWPTAPCLCNA0NCNTTNTNAIGNTP0RDAYGNGM0GNS0FOQ 289
QY 200 ANEYRPHNG--000SNRYGR0Y0GGN0NSNG0HESFNSFNFSQ0NNN 247
Db 290 AN0N00Y0NT0Y000NG0NG0N0F--00SGN0N0NS0Y00TSM00MNN 335

```

RESULT 4  
T20847  
hypothetical protein F13E9.4 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T20847  
R:McMurray, A.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z19332  
A:Accession: T20847  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-409 <MWL>  
A:Cross-references: EMBL:Z69383; PIDN:CAA93412.1; GSPDB:GN00022; CESP:F13E9.4  
C:Genetics:  
A:Experimental source: clone F13E9  
A:Gene: CESP:F13E9.4  
A:Map position: 4  
A:Introns: 32/1; 275/3; 337/3

Query Match	17.2%	Score 232	DB 2	Length 409
Best Local Similarity	27.9%	Pred. No. 6	9e-09	
Matches	93	Conservative	29	Mismatches 121, Indels 90, Gaps 15

  

QY	1	OGGOGGOGGOG	--OGG	-----	OGGSGFTALSLASFEINNNNOG	---Q	43	
DB	66	OGYOGNOGSMQY	SOOGYGGNSQODYGY	SOSGSGMGYGYGGS	--SGSYGQAFACQ		122	
QY	44	NQSGGSGFGALASMA	SFPMHS	-----	NNNNNSNSOGYNOGYON	-----	NONS 89	
DB	123	RPOGFQSGFSGGQA	SGSMNSFGG	GGGAGYGQ	NONGFGGSGF	--SGSGGMSLSLSSANS	180	
QY	90	OGYNNQOYOGGNGY	---QQGSGGAFSSLASM	-----	-----	-AOSYLG	127	
DB	181	NGNNN	---GSSSGYQNNQGR	HOQGGGGGSSSSSYMSNNGY	SSNSGYNNGN	PTPSFLN	237	
QY	128	GGGTOSNQOY	-----	NQGGNNQOY	QOQOQY	OHQOQO	---QQOQGH	170
DB	238	NVSSSAADPYNI	VANKSLTTNQIN	EQASNMASANSYQ	MYIYQYENR	SAIHTQATONS	297	

```

OY      1  0G0GGG0G0G0G---0GGG-----0G0GFTALASLSEFMSNNNNNOG-----Q 43
Db      66  0G0G0N0GSM0GYS0QGTGGGNSQDDYGTSG0GSGMGV0GTGGG---SGTSG00AF0AQQ 122
OY      44  N0SGGSGF0ALASMASSFMHS-----NN0N0SNN0GQYNOGTY0NG-----N0NS 89
Db      123 R0P0GFG0GNGFSG0Q0AS0MSHSFG0G0GT0G0N0G0FG0QSGF--SG0GMS0NLS0SSNS 180
OY      90  0GYNN0QY0GGNGTY-----0Q0G0SGC0AFBSLASM-----A0SYLG 127
Db      181 NGNNN---0GSSSGY0NN0G0RH0G0GGGSHSSSSNSVMSNN0YSNSGYNNGNPTPSEFLN 237
OY      128  G0G0T0SNOQY-----N0G0NN0Q0Y0Q0G0NY0HQ0Q0-----0Q0GHS 176
Db      238 NVSS0A0DY0N1YVNNKSJLTJN0INE0ASMA0NSV0A0I0Y0ETNRSALH0T0AT0NV5 297

```











GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2002, 15:02:16 ; Search time 19.74 Seconds  
(without alignments)  
288.416 Million cell updates/sec

Title: US-09-591-632-50\_COPY\_153\_405

Perfect score: 1347  
Sequence: 1 OGCGGCGGCGGCGGCGGCGG.....SFNFGNFGQNNNGNNGNRY 253

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	198	14.7	678	5 PCT-US93-03027-3	Sequence 3, Appl
2	183.5	13.6	2703	1 US-08-185-432-19	Sequence 19, Appl
3	163.5	12.1	595	1 US-08-425-069-4	Sequence 4, Appl
4	163.5	12.1	595	2 US-08-317-844B-4	Sequence 4, Appl
5	158.5	11.8	903	2 US-08-853-310-2	Sequence 2, Appl
6	156.5	11.6	401	4 US-09-219-849-34	Sequence 34, Appl
7	155.5	11.5	529	4 US-09-247-806-2	Sequence 2, Appl
8	150	11.1	485	2 US-08-749-391-2	Sequence 2, Appl
9	150	11.1	485	2 US-09-390-200-2	Sequence 2, Appl
10	145	10.8	542	1 US-07-814-964-13	Sequence 13, Appl
11	145	10.8	542	1 US-08-258-442-13	Sequence 13, Appl
12	145	10.8	542	1 US-08-328-809-8	Sequence 13, Appl
13	145	10.8	542	5 PCT-US92-11107-13	Sequence 13, Appl
14	145	10.6	729	4 US-09-625-188-20	Sequence 20, Appl
15	142.5	10.6	493	4 US-08-556-978B-59	Sequence 59, Appl
16	141.5	10.5	714	4 US-08-556-978B-61	Sequence 61, Appl
17	141.5	10.5	714	4 US-09-247-806-10	Sequence 10, Appl
18	138.5	10.3	760	1 US-08-195-152-2	Sequence 2, Appl
19	131	9.7	668	1 US-08-468-036-5	Sequence 5, Appl
20	131	9.7	668	2 US-08-376-843-5	Sequence 5, Appl
21	130.5	9.7	431	4 US-09-416-050A-8	Sequence 8, Appl
22	130.5	9.7	431	4 US-09-664-800-8	Sequence 8, Appl
23	130.5	9.7	431	4 US-09-665-309-8	Sequence 8, Appl
24	130.5	9.7	431	4 US-09-661-569-8	Sequence 8, Appl
25	129	9.6	651	4 US-08-556-978B-19	Sequence 19, Appl
26	129	9.6	651	4 US-09-247-806-1	Sequence 1, Appl
27	129	9.6	718	1 US-08-425-069-2	Sequence 2, Appl

28	129	9.6	718	2 US-08-317-844B-2	Sequence 2, Appl
29	129	9.6	747	3 US-09-034-177-3	Sequence 3, Appl
30	126.5	9.4	341	2 US-08-538-711A-8	Sequence 8, Appl
31	126.5	9.4	341	4 US-08-725-027-8	Sequence 8, Appl
32	126.5	9.4	353	4 US-08-538-711A-7	Sequence 7, Appl
33	126.5	9.4	353	4 US-08-725-027-7	Sequence 7, Appl
34	126.5	9.4	414	1 US-08-343-682-1	Sequence 1, Appl
35	126.5	9.4	414	2 US-08-705-660-26	Sequence 26, Appl
36	126.5	9.4	414	3 US-08-989-045-26	Sequence 26, Appl
37	126	9.4	591	3 US-08-965-903B-2	Sequence 2, Appl
38	125.5	9.3	428	1 US-08-190-802A-29	Sequence 29, Appl
39	125.5	9.3	428	4 US-08-477-346-29	Sequence 29, Appl
40	125	9.3	656	2 US-08-343-443B-2	Sequence 2, Appl
41	125	9.3	656	4 US-09-214-564A-4	Sequence 4, Appl
42	125	9.3	1250	4 US-08-441-139-9	Sequence 9, Appl
43	123	9.1	606	4 US-08-556-978B-21	Sequence 21, Appl
44	123	9.1	606	4 US-09-247-806-4	Sequence 4, Appl
45	122.5	9.1	353	2 US-08-687-702-37	Sequence 37, Appl

#### ALIGNMENTS

RESULT 1  
Sequence 3, Application PC/TUS9303027  
GENERAL INFORMATION:  
APPLICANT: LEONARD, WARREN; TOLEDANO,  
MICHEL  
TITLE OF INVENTION: CONTROL AND/OR  
PREVENTION OF BINDING OF NF-B/REL/DORSAL  
TITLE OF INVENTION:  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/03027  
FILING DATE: 19930401  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/862,987  
FILING DATE: 06-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: DOROTHY R. AUTH  
REGISTRATION NUMBER: P-36,434  
REFERENCE/DOCKET NUMBER: 2026-4010 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-758-4800  
TELEFAX: 212-751-6849  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 678  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Drosophila melanogaster  
STRAIN: Oregon R  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE: embryo  
HAPLOTYPE:  
TISSUE TYPE:







Sequence 2, Application US/08749391  
Patent No. 5948667  
GENERAL INFORMATION:  
APPLICANT: Cheng, Kuo-Joan  
APPLICANT: Selinger, Leonard B.  
APPLICANT: Liu, Jin-Hao  
APPLICANT: Hu, Youji  
APPLICANT: Forsberg, Cecil W.  
APPLICANT: Moloney, Maurice M.  
TITLE OF INVENTION: A xylanase Obtained From an  
TITLE OF INVENTION: Anaerobic Fungus  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80803  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/749,391  
FILING DATE: 13-NOV-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Donna M. Ferber  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 93-96  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 485 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-749-391-2

Query Match 11.1%; Score 150; DB 2; Length 485;  
Best Local Similarity 28.4%; Pred. No. 7.4e-06;  
Matches 69; Conservative 11; Mismatches 117; Indels 46; Gaps 8;  
QY 14 GCGCGGCTALASLASFFMNSNNNNQGNQSSGSSFGALASMASSFFMHSNNNNNSN 73  
DB 223 GNGNGVSGTADPPYAKVYIGDNGGASPAAPAGAPAGAPA-----GND 268  
QY 74 SGGYNQSYNGNNGNSGYNNOYQGGNGGYQDQGGSGGAFSSLASMAOSYLGCGTQS 133  
DB 269 QPQG-----PQGGQPPQGQPPQGQPPQGQPPQGQPPQGQ-----PPQG 305  
QY 134 NQOQYNO--QGQNNQ--QYQDQGNQYHQDQGGQDQGGHSSFSALASMASSYLGNNNS 190  
DB 306 NDDQGGQPPQGQPPQGNDQDQGGQ--PPQPPQGPPGGSDFFNNNNQGGSPWGGN--QG 362  
QY 191 NSSYGGQQAQNEYGPRQHNQDQGSNEXGRPOYGGNNNSN--GQHSFNFSGNFSQNNNN 247  
DB 363 GSPWGGNNGNPNMGNGGSSPWWGNGGSPWGGNNGGNNMGNGGSPWGGNNGGNNPWW 422  
QY 248 GNG 250  
DB 423 GNG 425

RESULT 9  
US-09-390-200-2  
Sequence 2, Application US/09390200  
Patent No. 6137032

GENERAL INFORMATION:  
APPLICANT: Cheng, Kuo-Joan  
APPLICANT: Selinger, Leonard B.  
APPLICANT: Liu, Jin-Hao  
APPLICANT: Hu, Youji  
APPLICANT: Forsberg, Cecil W.  
APPLICANT: Moloney, Maurice M.  
TITLE OF INVENTION: A xylanase Obtained From an  
TITLE OF INVENTION: Anaerobic Fungus  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80803  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/390,200  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/749,391  
FILING DATE: 13-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Donna M. Ferber  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 93-96  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 485 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-390-200-2

Query Match 11.1%; Score 150; DB 4; Length 485;  
Best Local Similarity 28.4%; Pred. No. 7.4e-06;  
Matches 69; Conservative 11; Mismatches 117; Indels 46; Gaps 8;  
QY 14 GCGCGGCTALASLASFFMNSNNNNQGNQSSGSSFGALASMASSFFMHSNNNNNSN 73  
DB 223 GNGNGVSGTADPPYAKVYIGDNGGASPAAPAGAPAGAPA-----GND 268  
QY 74 SGGYNQSYNGNNGNSGYNNOYQGGNGGYQDQGGSGGAFSSLASMAOSYLGCGTQS 133  
DB 269 QPQG-----PQGGQPPQGQPPQGQPPQGQPPQGQPPQGQ-----PPQG 305  
QY 134 NQOQYNO--QGQNNQ--QYQDQGNQYHQDQGGQDQGGHSSFSALASMASSYLGNNNS 190  
DB 306 NDDQGGQPPQGQPPQGNDQDQGGQ--PPQPPQGPPGGSDFFNNNNQGGSPWGGN--QG 362  
QY 191 NSSYGGQQAQNEYGPRQHNQDQGSNEXGRPOYGGNNNSN--GQHSFNFSGNFSQNNNN 247  
DB 363 GSPWGGNNGNPNMGNGGSSPWWGNGGSPWGGNNGGNNMGNGGSPWGGNNGGNNPWW 422  
QY 248 GNG 250  
DB 423 GNG 425

RESULT 10  
US-07-814-964-13  
Sequence 13, Application US/07814964





[illegible]

```

: RESULT 14
: US-09-625-188-20
: Sequence 20, Application US/09625188
: Patent No. 6307037
: GENERAL INFORMATION:
: APPLICANT: No. 6307037artis AG
: TITLE OF INVENTION: Fungal Target Genes and Methods
: FILE REFERENCE: PB/5-31285P1
: CURRENT APPLICATION NUMBER: US/09/625,188
: CURRENT FILING DATE: 2000-07-21
: NUMBER OF SEQ ID NOS: 44
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 20
: LENGTH: 729
: TYPE: PRF
: ORGANISM: Ashyba gossypii
: US-09-625-188-20

```

[illegible]

RESULT 15  
US-08-556-978B-59  
Sequence 59, Application US/08556978B  
Patent No. 6268169  
GENERAL INFORMATION:  
APPLICANT: FAHNESTOCK, STEPHEN F.  
TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED  
TITLE OF INVENTION: SPIDER SILK ANALOGS

1 NUMBER OF SEQUENCES: 107  
 2 CORRESPONDENCE ADDRESS:  
 3 ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
 4 STREET: 1007 MARKET STREET  
 5 CITY: WILMINTON  
 6 STATE: DELAWARE  
 7 COUNTRY: UNITED STATES OF AMERICA  
 8 ZIP: 19898  
 9  
 10 COMPUTER READABLE FORM:  
 11 MEDIUM TYPE: DISKETTE, 3.50 INCH  
 12 COMPUTER: IBM PC COMPATIBLE  
 13 OPERATING SYSTEM: MICROSOFT WINDOWS 95  
 14 SOFTWARE: MICROSOFT WORD FOR WINDOWS 95  
 15 CURRENT APPLICATION DATA:  
 16 APPLICATION NUMBER: US/08/556,978B  
 17 FILING DATE:  
 18 CLASSIFICATION: 435  
 19 PRIOR APPLICATION DATA:  
 20 APPLICATION NUMBER: 08/077,600  
 21 FILING DATE: JUNE 15, 1993  
 22 ATTORNEY/AGENT INFORMATION:  
 23 NAME: FLOYD, LINDA AXAMETHY  
 24 REGISTRATION NUMBER: 33,692  
 25 REFERENCE/DOCKET NUMBER: CR-9389-A  
 26 TELECOMMUNICATION INFORMATION:  
 27 TELEPHONE: 302-892-8112  
 28 TELEFAX: 302-773-0164  
 29 INFORMATION FOR SEQ ID NO: 59:  
 30 SEQUENCE CHARACTERISTICS:  
 31 LENGTH: 493 amino acids  
 32 TYPE: amino acid  
 33 STRANDEDNESS: unknown  
 34 TOPOLOGY: unknown  
 35 MOLECULE TYPE: protein  
 36 US-08-556-978B-59

[illegible]

Search completed: February 1, 2002, 15:02:17  
Job time: 237 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: February 1, 2002, 15:01:49 ; Search time 38.77 Seconds  
(without alignments)  
483.377 Million cell updates/sec

Title: US-09-591-632-50\_COPY\_153\_405  
Perfect score: 1347  
Sequence: 1 QGQGGGQGGGQGGGQGGG.....SFNFSGNFSQNNNGNNGNRY 253

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_1101:\*

- 1: /SIDSB/gcgdata/geneseq/geneseqp/AA1980.DAT.\*
- 2: /SIDSB/gcgdata/geneseq/geneseqp/AA1981.DAT.\*
- 3: /SIDSB/gcgdata/geneseq/geneseqp/AA1982.DAT.\*
- 4: /SIDSB/gcgdata/geneseq/geneseqp/AA1983.DAT.\*
- 5: /SIDSB/gcgdata/geneseq/geneseqp/AA1984.DAT.\*
- 6: /SIDSB/gcgdata/geneseq/geneseqp/AA1985.DAT.\*
- 7: /SIDSB/gcgdata/geneseq/geneseqp/AA1986.DAT.\*
- 8: /SIDSB/gcgdata/geneseq/geneseqp/AA1987.DAT.\*
- 9: /SIDSB/gcgdata/geneseq/geneseqp/AA1988.DAT.\*
- 10: /SIDSB/gcgdata/geneseq/geneseqp/AA1989.DAT.\*
- 11: /SIDSB/gcgdata/geneseq/geneseqp/AA1990.DAT.\*
- 12: /SIDSB/gcgdata/geneseq/geneseqp/AA1991.DAT.\*
- 13: /SIDSB/gcgdata/geneseq/geneseqp/AA1992.DAT.\*
- 14: /SIDSB/gcgdata/geneseq/geneseqp/AA1993.DAT.\*
- 15: /SIDSB/gcgdata/geneseq/geneseqp/AA1994.DAT.\*
- 16: /SIDSB/gcgdata/geneseq/geneseqp/AA1995.DAT.\*
- 17: /SIDSB/gcgdata/geneseq/geneseqp/AA1996.DAT.\*
- 18: /SIDSB/gcgdata/geneseq/geneseqp/AA1997.DAT.\*
- 19: /SIDSB/gcgdata/geneseq/geneseqp/AA1998.DAT.\*
- 20: /SIDSB/gcgdata/geneseq/geneseqp/AA1999.DAT.\*
- 21: /SIDSB/gcgdata/geneseq/geneseqp/AA2000.DAT.\*
- 22: /SIDSB/gcgdata/geneseq/geneseqp/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1347	100.0	405	22	AAB30810
2	1347	100.0	405	22	AAB30821
3	895	66.4	407	10	AAP94879
4	218	16.2	954	21	AAB48550
5	203.5	15.1	461	16	AAB75506
6	198.5	14.7	280	21	AAB63197
7	198	14.7	678	17	AAR42087
8	192	14.3	342	14	AAW03625
9	189.5	14.1	145	22	AAB69614
10	187.5	13.9	894	22	AAB35148
11	184.5	13.7	271	22	AAB30800

12	184.5	13.7	437	16	AAR75507
13	184.5	13.7	729	22	AAG70847
14	183.5	13.6	590	12	AAR11019
15	179.5	13.3	155	22	AAB69612
16	178	13.2	113	22	AAB69615
17	175	13.0	256	13	AAR24796
18	175	13.0	724	21	AG31571
19	175	13.0	826	21	AG31570
20	173.5	12.9	913	21	AG31569
21	173.5	12.9	404	17	AAR90675
22	171.5	12.7	326	21	AG37007
23	171.5	12.7	342	21	AG37005
24	171.5	12.7	343	21	AG37005
25	169.5	12.6	149	22	AAB72673
26	169	12.5	537	7	AAPE0452
27	168.5	12.5	685	22	AAB30792
28	165	12.2	477	22	AAB30815
29	164	12.2	715	22	AAB30820
30	163.5	12.1	531	20	AAI40098
31	163.5	12.1	595	12	AAR14309
32	163.5	12.1	595	12	AAW53347
33	163.5	12.1	595	19	AAV59071
34	162.5	12.1	767	11	AAR08401
35	162.5	12.1	767	11	AAR07599
36	162.5	12.1	2074	21	AAI54319
37	161.5	12.0	121	22	AAB69609
38	158.5	11.8	903	18	AAW37389
39	158.5	11.8	903	18	AAW37391
40	156.5	11.6	401	21	AAB06023
41	156.5	11.6	599	22	AAI72374
42	156.5	11.6	599	22	AAI72375
43	155.5	11.5	1844	21	AAI18250
44	155.5	11.5	3190	22	AAB84634
45	154.5	11.5	409	22	AAG70829

#### ALIGNMENTS

RESULT 1	
AA30810	
ID AAB30810 standard: Protein: 405 AA.	
XX	
AC AAB30810:	
XX	
DT 02-APR-2001 (first entry)	
XX	
DE Amino acid sequence of a prion-like amyloidogenic protein.	
XX	
KW SCHAG; self-coalesce; higher-order aggregate; amyloidogenic domain;	
KW aggregation; fibril; phenotypic alteration; gene therapy;	
KW disease resistance; plant pigmentation; prion disease.	
XX	
OS Saccharomyces cerevisiae.	
XX	
PN WO200075324-A2.	
XX	
PD 14-DEC-2000.	
XX	
PE 09-JUN-2000: 2000WO-US15876.	
XX	
PR 09-JUN-1999: 99US-0138833.	
XX	
PA (ARCH-) ARCH DEV CORP.	
XX	
PI Lindquist S, Li L, Ma J, Liu J, Sondheimer N, Scheibel T;	
XX	
DR WPI: 2001-061723/07.	
XX	
PT New nucleic acid encoding chimeric proteins with self-assembly	
PT properties, useful e.g. for diagnosis and treatment of prion diseases,	
XX	
PT also related aggregates, fibrils and polymers -	
XX	

Nicotiana plumbagi  
C albicans apoptos  
Apo-11proteins E  
Huntingtin accuml  
Sequence of fibron  
Arabidopsis thaila  
Arabidopsis thaila  
Arabidopsis thaila  
Human chromosome-1  
Arabidopsis thaila  
Arabidopsis thaila  
Polyglutamine trac  
Sequence of the As  
Amino acid sequenc  
Amino acid sequenc  
Amino acid sequenc  
Spider silk protei  
N.clavipes draglin  
Nephila clavipes s  
N. clavipes spider  
Amf105 encoded by  
Sequence deduced F  
Amino acid sequenc  
Huntingtin accuml  
Human additional s  
Human additional s  
Polar gelatin P te  
Amphiphilic recomb  
Amphiphilic recomb  
Plasmodium falcipa  
Amino acid sequenc  
C albicans apoptos











Db 700 -----sgggggggsgsgsgyyqgdhynspvpkagkxkphgqgkpsysgy 749

Qy 105 000GSGGSAFSLASMAQS-YLGGGQTQSNQOQYN--OQGNQOQYQOQ-----GQNYQ 157

Db 750 gahgqgq-----qsyngspysnygppgkqkqyghgsgysysnsyssp9gsgsdydn 802

Qy 158 HQOQGCQOQOQHSSSFSALASMASSYLGNNSNSNSSYGGOQOQANEXGRPOHNGOQOQSNRY 217

Db 803 yeaefnygsgrsgsgsgnygsqgaay---npqshggygsgsgsgsgsdyqgkqgysqsn-y 858

Qy 218 GRPOYGGNONSNGQHESENFES-GNFSQOQNNNGNQRX 253

Db 859 nsp--gsqgnygsppssygsqsggygrnadhemygy 893

RESULT 11

AAB30800

ID AAB30800 standard; Protein; 271 AA.

XX AAB30800;

AC

XX 02-APR-2001 (first entry)

DE A modified N region of yeast Sup35 protein.

XX

KW Sup35; phenotypic; SCHAG; self-coalesce; higher-order aggregate;

KW amyloidogenic domain; aggregation; fibril; phenotypic alteration;

KW gene therapy; disease resistance; plant pigmentation; prion disease.

XX

OS Synthetic.

OS Saccharomyces cerevisiae.

XX

PN WO200075324-A2.

XX

PD 14-DEC-2000.

XX

PF 09-JUN-2000; 2000WO-US15876.

XX

PR 09-JUN-1999; 99US-0138833.

PA (ARCH-) ARCH DEV CORP.

PI Lindquist S, Li L, Ma J, Liu J, Sondheimer N, Scheibel T;

XX

DR MPI: 2001-061723/07.

XX

DR N-PSDB: AAC86685.

XX

PT New nucleic acid encoding chimeric proteins with self-assembly

PT properties, useful e.g. for diagnosis and treatment of prion diseases,

PT also related aggregates, fibrils and polymers -

XX

PS Claim 11: Page 135-136; 188pp; English.

XX

CC The present sequence represents a modified N region of Sup35 protein, in

CC the second oligopeptide repeat has been expanded twice, creating a total

CC of seven repeats. Sup35 possesses the prion-like capacity to undergo a

CC self-perpetuating conformational alteration that changes the functional

CC state of Sup35 in a manner that creates a heritable change in phenotype.

CC It is used to construct chimeric polypeptides of the invention, which

CC comprise at least one SCHAG (self-coalesces into higher-order aggregates)

CC amino acid sequence fused in frame with a polypeptide of interest (which

CC is other than a marker protein, a glutathione-S-transferase or a

CC staphylococcal nuclear protein). The specification also describes

CC chimeric polypeptides that comprises an amyloidogenic domain that causes

CC aggregation into fibrils. The chimeric polypeptides are used to prepare

CC polymers with multiple reactivities, e.g. derivatised with enzymes, or

CC specific binding partners, and useful e.g. for performing multi-step

CC chemical reactions. They can be used create an inducible, or stable

CC phenotypic alteration in a cell, e.g. for gene therapy, protein

CC production, imparting disease resistance to plants, altering plant

CC pigmentation and for diagnosis and treatment of prion diseases.

XX

SQ Sequence 271 AA:

Query Match 13.7%; Score 184.5; DB 22; Length 271;

Best Local Similarity 31.3%; Pred. No. 5.8e-08;

Matches 63; Conservative 13; Mismatches 56; Indels 69; Gaps 10;

Qy 67 NNQNSNNSOQGYNQSYONGQONS-----QSYN--NOQYGCNGCYOQOQSGGSAFSLA 119

Db 2 sdsngnngnqnyqyqgngnqgqgnrryqgyqaynaqap2ggyyqnyqy9g----- 54

Qy 120 SMAQSLGSGGQTQSNQOQYNQOQGNQOQYQOQGNQYQHQQGCGOQOQCHSSSFSALASW 179

Db 55 -----Yqggg-----Yqgyypqg--Ygyqyppggyqgympdagyqqq----- 90

Qy 180 ASSYLGNNSNSNSSYGGOQOQANEXGRPOHNGOQOQSNRYGRPOYGGNONSNGQHESENFSG 239

Db 91 -----Ynpqggyqqyn-----pqggyqqqfn-----pqgg-----rgnyknfnym 126

Qy 240 N-----FSOQNNNGNQRX 253

Db 127 nlqgyqagfgpqsqgmIndf 147

RESULT 12

AAR75507

ID AAR75507 standard; Protein; 437 AA.

XX AAR75507;

AC

XX 05-DEC-1995 (first entry)

DE Nicotiana plumbaginifolia arabinogalactan protein NPACP1.

XX

KW Arabinogalactan protein; AGP; gum; thickening agent; emulsifier;

KW adhesive.

XX

OS Nicotiana plumbaginifolia.

XX

FH Key Location/Qualifiers

FT Peptide 14..24

FT Peptide /note="corresp. to microsequenced peptide"

FT Peptide 14..24

FT /note="Identical with microsequenced peptide"

FT MISC-difference 28

FT /note="see above"

FT MISC-difference 30

FT /note="see above"

FT MISC-difference 32..37

FT /note="see above"

FT MISC-difference 25

FT /label="hydroxyproline

FT MISC-difference 27

FT /label="see above

FT MISC-difference 29

FT /label="see above

FT MISC-difference 30

FT /label="see above

FT MISC-difference 26

FT /note="T in microsequenced peptide"

PN WO9515377-A.

PD 08-JUN-1995.

XX

PF 01-DEC-1994; 94WO-AU00744.

XX

PR 18-JUL-1994; 94US-0276452.

XX

PR 03-DEC-1993; 93US-0161944.

XX

PA (ALBR ) ALBRIGHT & WILSON AUSTRALIA.

PI Bacic A, Chen C, Clarke AE, Du H, Gane AM, Mau S;









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2002, 15:01:55 ; Search time 37.66 Seconds

(without alignments)  
477.735 Million cell updates/sec

Title: US-09-591-632-2\_COPY\_1\_123

Perfect score: 701  
Sequence: 1 MSDSNOGNQONTQOVSQNG.....FVNNNLQGYQAGFPQPSOG 123

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: SP:REMBL\_17:\*  
2: SP:archaea:\*  
3: SP:bacteria:\*  
4: SP:fungi:\*  
5: SP:human:\*  
6: SP:mammal:\*  
7: SP:invertebrate:\*  
8: SP:mhc:\*  
9: SP:organelle:\*  
10: SP:phage:\*  
11: SP:plant:\*  
12: SP:rodent:\*  
13: SP:virus:\*  
14: SP:vertebrate:\*  
15: SP:unclassified:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	701	100.0	685	3	Q9HGV1
2	695	99.1	224	3	Q07089
3	229.5	32.7	712	3	Q9HGI7
4	228	32.5	299	3	Q9UVK3
5	212	30.2	428	3	Q9URS3
6	204	29.1	700	3	Q9HGI8
7	194.5	27.7	701	3	Q9HGI6
8	190	25.2	662	3	Q9HGI4
9	177	25.0	742	3	Q9HGI5
10	175	25.0	742	3	Q9HGI5
11	170	24.3	200	11	Q9D036
12	167.5	23.9	305	3	Q9UVK0
13	162	23.1	2382	5	Q9B119
14	160	22.8	407	3	Q94079
15	152.5	21.8	215	3	Q43117
16	149	21.3	456	5	Q20936
17	148	21.1	419	10	Q49429
18	147.5	21.0	409	5	Q19414

20	147.5	21.0	924	5	Q20189	Q20189 caenorhabd1
21	146	20.8	497	5	Q27512	Q27512 caenorhabd1
22	144	20.5	232	5	Q20013	Q20013 caenorhabd1
23	143	20.4	584	5	Q9NJ80	Q9NJ80 spodoptera
24	141.5	20.2	492	4	Q9NRD7	Q9NRD7 homo sapien
25	141	20.1	800	13	Q91551	Q91551 xenopus lae
26	140.5	20.0	236	3	Q94060	Q94060 candida alb
27	139.5	19.9	422	4	Q9BXC6	Q9BXC6 homo sapien
28	139.5	19.9	492	4	Q9HBC2	Q9HBC2 homo sapien
29	139.5	19.9	716	3	Q42787	Q42787 podospora a
30	139	19.8	632	10	Q82762	Q82762 arabidopsis
31	138	19.7	1111	2	Q86522	Q86522 streptomyces
32	137.5	19.6	284	5	Q21073	Q21073 caenorhabd1
33	137.5	19.6	765	5	Q17796	Q17796 caenorhabd1
34	136.5	19.5	461	10	Q41256	Q41256 nicotiana a
35	135	19.3	528	4	Q13344	Q13344 homo sapien
36	135	19.3	867	5	Q18218	Q18218 caenorhabd1
37	135	19.3	871	5	Q62486	Q62486 caenorhabd1
38	134.5	19.2	319	10	Q41725	Q41725 zinnia eleg
39	133.5	19.0	254	5	Q9U7P0	Q9U7P0 eufoiliacul1
40	133	19.0	369	4	Q9HBC1	Q9HBC1 homo sapien
41	133	19.0	853	5	Q94464	Q94464 dictyostell
42	133	19.0	3498	5	Q20497	Q20497 caenorhabd1
43	131	18.7	1118	5	Q9TYH5	Q9TYH5 plasmodium
44	131	18.7	1161	5	Q9TY17	Q9TY17 plasmodium
45	131	18.7	1961	3	Q9UVK9	Q9UVK9 yarrowia 11

#### ALIGNMENTS

RESULT 1  
Q9HGV1 PRELIMINARY: PRT: 685 AA.  
AC Q9HGV1;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE SUP35 ALLOSUPPRESSOR MUTANT SAL3-4.  
GN SUP35.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BSC783/4A SAL3-4;  
RA Resende C.G., Duarte J.B., Tuile M.F.;  
RT "sal3-4 a allosuppressor mutation of the SUP35 gene.";  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.  
DR EMBL: AF263099; AAF9684.1; -;  
DR InterPro: IPR000795; GTP\_EFTU.  
DR Pfam: PF00009; GTP\_EFTU; 1.  
DR PRINTS: PR00315; ELONGATNCT.  
KW GTP-binding; Protein biosynthesis.  
SQ SEQUENCE 685 AA; 76550 MW; F3958EC37D750154 CRC64;

Query Match 100.0%; Score 701; DB 3; Length 685;  
Best Local Similarity 100.0%; Pred. No. 9.3e-61;  
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSDSNOGNQONTQOVSQNGNNRQGYQAYNQAOPAGGYONYOQSGYQOOGY 60  
Db 1 MSDSNOGNQONTQOVSQNGNNRQGYQAYNQAOPAGGYONYOQSGYQOOGY 60  
QY 61 QQYNPDAGYQOQYNNPQGYQOQYNNPQGYQOQYNNPQGYQOQYNNPQGYQOQY 120  
Db 61 QQYNPDAGYQOQYNNPQGYQOQYNNPQGYQOQYNNPQGYQOQYNNPQGYQOQY 120  
QY 121 SOG 123  
111

DB 121 SQ 123

RESULT 2

007089 PRELIMINARY: PRT: 224 AA.

AC 007089;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE OMNIPOTENT SUPPRESSOR.

GN SUP3 OR YDR172W.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=88329727; PubMed=3047009;

RA Kushnir V.V., Ter-Avanesyan M.D., Telckov M.V., Surguchov A.P.,

RT "Nucleotide sequence of the SUP2 (SUP35) gene of Saccharomyces

RT cerevisiae.";

RL Gene 66:45-54(1988).

CC -1- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.

DR EMBL; M21129; AAA5132.1; -.

DR SGD; S0002579; SUP35.

DR InterPro: IPR000795; GTP\_EFTU.

DR Pfam; PR00009; GTP\_EFTU; 1.

KW GTP-binding; Protein biosynthesis.

SQ SEQUENCE 224 AA; 25182 MW; 2521133DC4C77D65 CRC64;

Query Match 99.1%; Score 695; DB 3; Length 224;

Best Local Similarity 100.0%; Pred. No. 1.1e-60;

Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSDSNGNNOQNTQOYSONGNOQGNRRYQGYAYNNAQAQAPAGGYQYQYSGYQGGY 60

DB 1 MSDSNGNNOQNTQOYSONGNOQGNRRYQGYAYNNAQAQAPAGGYQYQYSGYQGGY 60

QY 61 QOYNPAGYQOQYNPQGGYQOQYNPQGGYQOQYNPQGGYQOQYNPQGGYQOQYNPQGGY 120

DB 61 QOYNPAGYQOQYNPQGGYQOQYNPQGGYQOQYNPQGGYQOQYNPQGGYQOQYNPQGGY 120

QY 121 SQ 122

DB 121 SQ 122

RESULT 3

09HG17 PRELIMINARY: PRT: 712 AA.

AC 09HG17;

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE POLYPEPTIDE RELEASE FACTOR 3.

GN SUP35.

OS Candida maltosa (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

CC Saccharomycetales; mitosporic Saccharomycetales; Candida.

OX NCBI\_TaxID=5479;

RN [1]

RP SEQUENCE FROM N.A.

RA Nakayashiki T., Edihara K., Nakamura Y.;

RT "Psi-like properties of fungus Sup35 proteins.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.

DR EMBL; AB039750; BAB12681.2; -.

DR InterPro: IPR000795; GTP\_EFTU.

DR Pfam; PR00009; GTP\_EFTU; 1.

DR PRINTS; PR00315; ELONGATNFCF.

KW GTP-binding; Protein biosynthesis.

SQ SEQUENCE 712 AA; 78631 MW; 729B6A814735C469 CRC64;

Query Match 32.7%; Score 229.5; DB 3; Length 712;

Best Local Similarity 48.8%; Pred. No. 1.1e-14;

Matches 60; Conservative 7; Mismatches 43; Indels 13; Gaps 7;

QY 4 SNGNNOQNTQOYSONGNOQGNRRYQGYAYNNAQAQAPAGGYQYQYSGYQGGYQ 61

DB 22 SKQPOQOQPOQOQYFNPQNAQAFVPTGGYQOQYQOQOQYQOQYQOQYQOQYQOQYQOQY 78

QY 62 Q-YNPAGYQOQYNPQGGYQOQYNPQGGYQOQYNPQGGYQOQYNPQGGYQOQYNPQGGY 119

DB 79 QYNNRGGYQOQYNNRGGYQOQYNNRGGYQOQYNNRGGYQOQYNNRGGYQOQYNNRGGY 132

QY 120 SQ 122

DB 133 QOQ 135

RESULT 4

09URK3 PRELIMINARY: PRT: 299 AA.

AC 09URK3;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

DE SUP35 HOMOLOG (FRAGMENT).

GN SUP35.

OS Candida albicans (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

CC Saccharomycetales; mitosporic Saccharomycetales; Candida.

OX NCBI\_TaxID=5476.

RN [1]

RP SEQUENCE FROM N.A.

RA Santoso A., Chien P., Weissman J.S.;

RT "Molecular Basis of a Yeast Prion Species Barrier.";

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF206287; AAF14002.1; -.

FT NON-TER 299 299

SQ SEQUENCE 299 AA; 32702 MW; F15D7800DE73C49F CRC64;

Query Match 32.5%; Score 228; DB 3; Length 299;

Best Local Similarity 43.2%; Pred. No. 6.2e-15;

Matches 57; Conservative 12; Mismatches 43; Indels 20; Gaps 8;

QY 1 MSDSNGNNOQNTQOYSONGNOQGNRRYQGYAYNNAQAQAPAGGYQYQYSGYQGGY 54

DB 15 MANASLNGDQSKQOQ 67

QY 55 YQO-GGYQOYNP-DAGYQOQYNPQGGYQOQYNPQGGYQOQYNPQGGYQOQYNPQGGY 111

DB 68 QOQYGGYQOQYNPQGGYQOQYNPQGGYQOQYNPQGGYQOQYNPQGGYQOQYNPQGGY 124

QY 112 GYQA-GRQPSQ 122

DB 125 GYQOYNSQPQOQ 136

RESULT 5

09URS3 PRELIMINARY: PRT: 428 AA.

AC 09URS3;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE SUP35 HOMOLOG (FRAGMENT).

GN SUP35.

OS Kluyveromyces lactis (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

CC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

OX NCBI\_TaxID=28985;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Santoso A., Chien P., Weissman J.S.:  
 RL "Molecular Basis of a Yeast Prion Species Barrier."  
 CC Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
 DE EMBL: AF206288; AAF14003.1; -  
 DR HSSP: P07157; IAIIP.  
 DR InterPro: IPR000795; GTP\_EFTU.  
 DR Pfam: PF00009; GTP\_EFTU.1.  
 DR PRINTS: PR00315; ELONGATNFACT.  
 DR PROSITE: PS00301; EFACITOR\_GTP.1.  
 KW GTP-binding; Protein biosynthesis.  
 FT NON\_TER 428  
 SQ SEQUENCE 428 AA: 47156 MW: 842508757D7DD6EF CRC64;

Query Match 30.2%; Score 212; DB 3; Length 428;  
 Best Local Similarity 47.0%; Pred. No. 3.3e-13;  
 Matches 62; Conservative 4; Mismatches 44; Indels 22; Gaps 10;

OY 1 MSD-SNQGNNQ-QNTQOYSQNGNQGGNNRYQGYQAYNAQAQAPAGGYQNYQGYSGYQOG 58  
 DB 1 MSDQNDQDGGQGYNQYNGQYNGYNGQ-QGYQNGQ-QGAPQGYQAYQAGQOPQG 58  
 OY 59 GYQOYNPDAGYQOQYNPQGGYQOQYNPQGGYQOQFN-PQGGRGYKKNFNYN-----NL 110  
 DB 59 AYQGYNPQA--QGYOP---YQGYNAQ---QCGYNAQGGHNNNKNKNNKSNYNNK 110  
 OY 111 QGYQA--GFQPG 120  
 DB 111 QGYQAGQYNAQ 122

RESULT 6  
 O9HG18 PRELIMINARY; PRT; 700 AA.  
 AC O9HG18;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE POLYPEPTIDE RELEASE FACTOR 3.  
 GN SUP35.  
 OS Kluyveromyces fragilis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OX NCBI\_TaxID=28985;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nakayashiki T., Edihara K., Nakamura Y.:  
 RL "Psi-like properties of fungus Sup35 proteins."  
 RT Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.  
 DR EMBL: AB039749; BAB12680.1; -  
 DR InterPro: IPR000795; GTP\_EFTU.  
 DR Pfam: PF00009; GTP\_EFTU.1.  
 DR PRINTS: PR00315; ELONGATNFACT.  
 KW GTP-binding; Protein biosynthesis.  
 SQ SEQUENCE 700 AA: 77639 MW: 53f9905442f6b025 CRC64;

Query Match 30.2%; Score 212; DB 3; Length 700;  
 Best Local Similarity 47.0%; Pred. No. 5.4e-13;  
 Matches 62; Conservative 4; Mismatches 44; Indels 22; Gaps 10;

OY 1 MSD-SNQGNNQ-QNTQOYSQNGNQGGNNRYQGYQAYNAQAQAPAGGYQNYQGYSGYQOG 58  
 DB 1 MSDQNDQDGGQGYNQYNGQYNGYNGQ-QGYQNGQ-QGAPQGYQAYQAGQOPQG 58  
 OY 59 GYQOYNPDAGYQOQYNPQGGYQOQYNPQGGYQOQFN-PQGGRGYKKNFNYN-----NL 110  
 DB 59 AYQGYNPQA--QGYOP---YQGYNAQ---QCGYNAQGGHNNNKNKNNKSNYNNK 110

OY 111 QGYQA--GFQPG 120  
 DB 111 QGYQAGQYNAQ 122

RESULT 7  
 O9UVK2 PRELIMINARY; PRT; 320 AA.  
 AC O9UVK2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE SUP35 HOMOLOG (FRAGMENT).  
 GN SUP35.  
 OS Kluyveromyces fragilis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OX NCBI\_TaxID=4911;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Santoso A., Chien P., Weissman J.S.:  
 RL "Molecular Basis of a Yeast Prion Species Barrier."  
 CC Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
 DE EMBL: AF206289; AAF14004.1; -  
 DR InterPro: IPR000795; GTP\_EFTU.  
 DR Pfam: PF00009; GTP\_EFTU.1.  
 KW GTP-binding; Protein biosynthesis.  
 FT NON\_TER 320  
 SQ SEQUENCE 320 AA: 35287 MW: 383023290245E051 CRC64;

Query Match 29.1%; Score 204; DB 3; Length 320;  
 Best Local Similarity 42.9%; Pred. No. 1.5e-12;  
 Matches 63; Conservative 5; Mismatches 45; Indels 34; Gaps 10;

OY 1 MSD-SNQGNNQ-QNTQOYSQNGNQGGNNRYQGYQAYNAQAQAPAGGYQNYQGYSGYQOG 58  
 DB 1 MSDQNDQDGGQGYNQYNGQYNGYNGQ-QGYQNGQ-QGAPQGYQAYQAGQOPQG 58  
 OY 59 GYQOYNPDAGYQOQYNPQGGYQOQYNPQGGYQOQFN-PQGGRGYKKNFNYN-----NL 110  
 DB 59 AYQGYNPQA--QGYOP---YQGYNAQ---QCGYNTQGGHNNNKNKNNKSNYNNK 110  
 OY 111 QGYQA--GFQPG 123  
 DB 111 QGYQAGQYNAQ 137

RESULT 8  
 O9HG16 PRELIMINARY; PRT; 701 AA.  
 AC O9HG16;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE POLYPEPTIDE RELEASE FACTOR 3.  
 GN SUP35.  
 OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Debaryomyces.  
 OX NCBI\_TaxID=4959;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nakayashiki T., Edihara K., Nakamura Y.:  
 RL "Psi-like properties of fungus Sup35 proteins."  
 RT Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.  
 DR EMBL: AB039751; BAB12682.2; -  
 DR InterPro: IPR000795; GTP\_EFTU.  
 DR Pfam: PF00009; GTP\_EFTU.1.  
 DR PRINTS: PR00315; ELONGATNFACT.





Db 1149 SQTVGFPQPPYSVKGYYQPHNQLNGYQNPQSSSFQA-NGYQPPSQ 1195

## RESULT 15

094079 ID 094079 PRELIMINARY; PRT; 407 AA.

AC 094079;

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE YEAST (S.CEREVISIAE) FUS1 (FUSION GENE) AND BIK1 GENE REGION.

GN RNO1 AND YCL028W.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=87286864; PubMed=3302672;

RA Trueheart J., Boeke J.D., Fink G.R.;

RT "Two genes required for cell fusion during yeast conjugation: evidence

for a pheromone-induced surface protein.";

RL MOL. CELL. BIOL. 7:2316-2328(1987).

DR EMBL; M16717; AAA34615.1; -.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

DR SGD; S0000533; RNO1.

Search completed: February 1, 2002, 15:07:10  
Job time: 315 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2002, 15:02:20 ; Search time 15.38 seconds  
(without alignments)  
293.223 Million cell updates/sec

Title: US-09-591-632-2\_COPY\_1\_123

Perfect score: 701  
Sequence: 1 MSDSNOGNNQNYQYQSNG.....FNYNNNLCGYQAGFQPSQG 123

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	701	100.0	685	1	ERF2_YEAST
2	228	32.5	715	1	ERF2_CANAL
3	167	23.8	592	1	R856_HUMAN
4	166	23.7	405	1	YCC8_YEAST
5	166	23.7	741	1	ERF2_PICPT
6	157	22.4	512	1	FUS_BOVIN
7	147	21.0	462	1	ANK7_DICDI
8	143	20.4	372	1	P65H_MYCGE
9	143	20.4	518	1	FUS_MOUSE
10	141	20.1	387	1	SSXT_MOUSE
11	139	19.8	418	1	SSXT_MOUSE
12	136	19.4	526	1	FUS_HUMAN
13	135	19.3	128	1	YBM6_YEAST
14	132	18.8	882	1	IF2_BORBU
15	128	18.3	353	1	ROD_RAT
16	127	18.1	355	1	ROD_HUMAN
17	126.5	18.0	738	1	YKRF_YEAST
18	123	17.5	346	1	YNS3_CAEEL
19	122.5	17.5	750	1	CBMB_BACTV
20	121	17.3	954	1	XYNA_RUMFL
21	118	16.8	838	1	GLT4_WHEAT
22	116	16.5	298	1	3AKD_MYCPA
23	115.5	16.5	826	1	CRMA_BACUH
24	115	16.4	839	1	GLT5_WHEAT
25	115	16.4	991	1	DHPI_SCHPO
26	114.5	16.3	552	1	ERF1_CANAL
27	114	16.3	295	1	LEG3_CANFA
28	113.5	16.2	307	1	SYPH_BOVIN
29	113	16.1	753	1	CKAA_BACUF
30	112	16.0	461	1	YS41_CAEEL
31	112	16.0	1028	1	OVO_CAEEL
32	111.5	15.9	313	1	SYPH_HUMAN
33	111	15.8	365	1	ROAL_DROME

34	111	15.8	870	1	YS05_CAEEL	Q10130 caenorhabd
35	110.5	15.8	346	1	RO21_XENLA	P51989 xenopus lae
36	110.5	15.8	1902	1	SMF1_HUMAN	O14497 homo sapien
37	109.5	15.6	159	1	CUP9_DROME	P27381 drosophila
38	107.5	15.3	1048	1	AC01_LARAT	O04379 arabidopsis
39	107	15.3	401	1	YK03_CAEEL	P34291 caenorhabd
40	107	15.3	1231	1	YK73_CAEEL	P34314 caenorhabd
41	106.5	15.2	660	1	GLT3_WHEAT	P08488 triticum ae
42	105.5	15.0	345	1	SOD_DROME	O08473 drosophila
43	105.5	15.0	648	1	GLT0_WHEAT	P10387 triticum ae
44	104.5	14.9	159	1	GRPA_MEDPA	O09134 medicago fa
45	104.5	14.9	655	1	EMS_MOUSE	O61545 mus musculu

## ALIGNMENTS

RESULT ID	ERF2_YEAST	STANDARD	PRT	685 AA
AC	P05453; P05420;			
DT	01-NOV-1988 (rel. 09, Created)			
DT	01-NOV-1988 (rel. 36, Last annotation update)			
DE	EUKARYOTIC PEPTIDE CHAIN RELEASE FACTOR GTP-BINDING SUBUNIT (ERF2)			
DE	(TRANSLATION RELEASE FACTOR 3) (ERF3) (ERF-3) (OMNIPOTENT SUPPRESSOR			
DE	PROTEIN 2) (G1 TO S PHASE TRANSITION PROTEIN 1)).			
GN	SUP35 OR SUP2 OR SUP12 OR GST1 OR SAL3 OR PNM2 OR YDR172W OR			
GN	YD9395.05.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RI	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88329727; PubMed=3047009;			
RA	Kushnirov V.V., Ter-Avanesyan M.D., Telckov M.V., Surguchov A.P.,			
RA	Smirnov V.N., Inge-Vechtomov S.G.;			
RT	"Nucleotide sequence of the SUP2 (SUP35) gene of Saccharomyces			
RT	cerevisiae.";			
RL	Gene 66:45-54(1988).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87219095; PubMed=3556215;			
RA	Kushnirov V.V., Ter-Avanesyan M.D., Surguchov A.P., Smirnov V.N.,			
RA	Inge-Vechtomov S.G.;			
RT	"Localization of possible functional domains in sup2 gene product of			
RT	the yeast Saccharomyces cerevisiae.";			
RL	FEBS Lett. 215:257-260(1987).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88172503; PubMed=3280807;			
RA	Wilson P.G., Culbertson M.R.;			
RT	"SUP2 suppressor protein of yeast. A fusion protein related to the			
RT	EF-1 family of elongation factors.";			
RL	J. Mol. Biol. 199:559-573(1988).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88296422; PubMed=2841115;			
RA	Kukuchi Y., Shmatake H., Kikuchi A.;			
RT	"A yeast gene required for the G1-to-S transition encodes a protein			
RT	containing an A-kinase target site and GTPase domain.";			
RL	EMBO J. 7:1175-1182(1988).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C / AB972;			
RA	Murphy L., Harris D.E., Barrell B.G., Rajandream M.A.;			
RL	Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.			
RN	[6]			
RP	FUNCTION.			
RX	MEDLINE=96016209; PubMed=7556078;			
RA	Stanfield I., Jones K.M., Kushnirov V.V., Dagkasmanakaya A.R.,			
RA	Poznyakovskii A.I., Pauskhin S.V., Ntarras C.R., Cox B.S.,			



```

RL Genomics 38:51-57(1996).
RN [2]
RP SEQUENCE FROM N.A. (SHORT ISOFORM), AND SEQUENCE OF 282-297 & 307-320.
RA MEDLINE-97045110; PubMed-8890175;
RX Bertolotti A., Lutz Y., Heard D.J., Chambon P., Tora L.;
RT "hNF(II)8, a novel RNA/ssDNA-binding protein with homology to the
RT pro-oncoproteins TLS/FUS and EMS is associated with both TFIIID and RNA
RT polymerase II.";
RL EMBO J. 15:5022-5031(1996).
RN [3]
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RX MEDLINE-99013873; PubMed-9795213;
RA Morchoshvili F., Ootsuka Y., Arai K., Ichikawa H., Mitani S.;
RA Munakata M., Ohki M.;
RT "Genomic structure of the human RBP56/hNAF168 and FUS/TLS genes.";
RL Gene 221:191-198(1998).
CC -1- FUNCTION: RNA AND SSDNA-BINDING PROTEIN THAT MAY PLAY SPECIFIC
CC ROLES DURING TRANSCRIPTION INITIATION AT DISTINCT PROMOTERS. CAN
CC ENTER THE PREINITIATION COMPLEX TOGETHER WITH THE RNA POLYMERASE
CC II (POL II).
CC -1- SUBUNIT: BELONGS TO THE RNA POLYMERASE II (POL II) TRANSCRIPTIONAL
CC MULTIPROTEIN COMPLEX, TOGETHER WITH THE TATA-BINDING PROTEIN (TBP)
CC AND OTHER TBP-ASSOCIATED FACTORS (TAF(II)S).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG ISOFORM (SHOWN HERE)
CC AND A SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBIQUITOUS. OBSERVED IN ALL FETAL AND ADULT
CC TISSUES.
CC -1- DISEASE: A FORM OF EXTRASKELETAL MYXOID CHONDROSARCOMAS (EMC) IS
CC CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(9;17)(Q22;Q11) THAT
CC INVOLVES TAF2N AND NR4A3.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -1- SIMILARITY: CONTAINS 1 RANBP-TYPE ZINC FINGER.
CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U51334; AAC05932.1; -
DR EMBL: X98893; CAA67398.1; -
DR EMBL: AB010067; BAA33811.1; -
DR EMBL: AB010057; BAA33811.1; JOINED.
DR EMBL: AB010058; BAA33811.1; JOINED.
DR EMBL: AB010059; BAA33811.1; JOINED.
DR EMBL: AB010060; BAA33811.1; JOINED.
DR EMBL: AB010061; BAA33811.1; JOINED.
DR EMBL: AB010062; BAA33811.1; JOINED.
DR EMBL: AB010063; BAA33811.1; JOINED.
DR EMBL: AB010064; BAA33811.1; JOINED.
DR EMBL: AB010065; BAA33811.1; JOINED.
DR EMBL: AB010066; BAA33811.1; JOINED.
DR EMBL: AB010067; BAA33812.1; -
DR EMBL: AB010057; BAA33812.1; JOINED.
DR EMBL: AB010058; BAA33812.1; JOINED.
DR EMBL: AB010059; BAA33812.1; JOINED.
DR EMBL: AB010060; BAA33812.1; JOINED.
DR EMBL: AB010061; BAA33812.1; JOINED.
DR EMBL: AB010062; BAA33812.1; JOINED.
DR EMBL: AB010063; BAA33812.1; JOINED.
DR EMBL: AB010064; BAA33812.1; JOINED.
DR EMBL: AB010065; BAA33812.1; JOINED.
DR EMBL: AB010066; BAA33812.1; JOINED.
DR MIM: 601574; -
DR InterPro: IPR000504; RRM.
DR InterPro: IPR001876; Znf-RanBP.
DR Pfam: PF00076; rrm; 1.
DR Pfam: PF00641; zf-RanBP; 1.
DR SMART: SM00360; RRM; 1.

```

```

DR SMART: SM00547; ZNF_RB2; 1.
DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
KW Proto-oncogene; RNA-binding; DNA-binding; Nuclear protein; Repeat;
KW Alternative splicing; Chromosomal translocation; Zinc-finger; Zinc;
KW Metal-binding.
FT DOMAIN 1 208 GIN/GLY/SER/TYR-RICH.
FT DOMAIN 234 320 RNA-BINDING (RRM).
FT DOMAIN 320 350 ARG/GLY-RICH.
FT ZN_FING 360 379 C4-TYPE (POTENTIAL).
FT DOMAIN 407 575 21 X APPROXIMATE/TANDEM REPEATS OF D-R-
FT [S,G](0,3)-G-G-Y-G-G.
FT REPEAT 407 413 1.
FT REPEAT 414 420 2.
FT REPEAT 421 429 3.
FT REPEAT 429 439 4.
FT REPEAT 430 439 5.
FT REPEAT 440 448 6.
FT REPEAT 449 457 7.
FT REPEAT 458 465 8.
FT REPEAT 466 473 9.
FT REPEAT 474 481 10.
FT REPEAT 482 488 11.
FT REPEAT 489 496 12.
FT REPEAT 497 503 13.
FT REPEAT 504 510 14.
FT REPEAT 511 517 15.
FT REPEAT 518 524 16.
FT REPEAT 525 533 17.
FT REPEAT 534 543 18.
FT REPEAT 544 551 19.
FT REPEAT 552 560 20.
FT REPEAT 561 568 21.
FT REPEAT 569 575 21.
FT VARSPLIC 60 62 MISSING (IN SHORT ISOFORM).
FT SEQUENCE 592 AA; 61830 MW; 73D37C171E1E2BCA CRC64;
SO
Query Match 23.8%; Score 167; DB 1; Length 592;
Best Local Similarity 29.1%; Pred. No. 8; 8e-07;
Matches 46; Conservative 17; Mismatches 45; Indels 50; Gaps 7;
QY 1 MSDS-----NGNNNOONQOYSONGNOQG-----NNRY-OGYOAYNAQAOPA 42
DB 1 MSDSGSGYSGSGEGQSQSTYTGPNQSGYGOASQSYSGYGTOTDSSYQNTSGYSSYQSQ 60
QY 43 GGYOYNOYGYSGYGOQGGYQO--YN-----PDAGYQOQYNPQG 77
DB 61 SGYSQSGYSGYENKQSSYSGQPYNNQGOQOQNMESGSGGGRAPSDQPDYGGQDSDYDQS 120
QY 78 GYOQYNPQGYQOQFNPQGRGNTKNTNNNNLQGYQA 115
DB 121 GYDOH--QGSYDEQSN-----YDOOHDSYSONOOSYHS 151
RESULT 4
YC8_YEAST STANDARD; PRT; 405 AA.
ID YC8_YEAST
AC P25367;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 42.6 KDA PROTEIN IN BIK1-FUS1 INTERGENIC REGION.
GN YCL028W OR YCL28W OR YCL181.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91377317; PubMed-1897318;
RA Red M.R., Luetzenkirchen K., Xu G., Kleinans U., Hollenberg C.P.;
RT "The complete sequence of a 11,953 bp fragment from Clg on chromosome
RT III encompasses four new open reading frames."

```





QY 108 NNLOQY--QAGFOPQ 120  
Db 110 ---QGYPPOQGYPPO 121

## RESULT 8

P65H\_MYCGE STANDARD; PRT; 372 AA.

DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PROLINE-RICH P65 PROTEIN HOMOLOG.

GN MG217

OS Mycoplasma genitalium.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;

OC Mycoplasmataceae; Mycoplasma.

OX NCBI\_TaxID=2097;

RN SEQUENCE FROM N.A.

RP STRAIN-ATCC 33530 / G-37;

RX MEDLINE=96026346; PubMed=7569993;

RA Fraser C.M., Gecayne J.D., White O., Adams M.D., Clayton R.A.,

RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,

RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,

RA Nguyen D.T., Uitterback T.R., Sauder D.M., Phillips C.A., Merrick J.M.,

RA Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,

RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,

RT "The minimal gene complement of Mycoplasma genitalium";

RL Science 270:397-403(1995).

CC -1- SIMILARITY: TO M.PNEUMONIAE PROTEIN P65.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: U39701; AAC71436.1; -

DR TIGR: MG217; -

KW Repeat; Complete proteome.

FT DOMAIN 29 128 8 X DPNOQ(O)FNQ REPEATS.

FT REPEAT 29 37 1.

FT REPEAT 41 49 2.

FT REPEAT 61 69 3.

FT REPEAT 81 89 4.

FT REPEAT 101 109 5.

FT REPEAT 120 128 6.

FT REPEAT 139 147 7.

FT REPEAT 120 128 8.

SO SEQUENCE 372 AA; 44664 MW; 4C29701D213CE19E CRC64;

Query Match 20.4%; Score 143; DB 1; Length 372;  
Best Local Similarity 32.7%; Pred. No. 4.8e-05;  
Matches 51; Conservative 12; Mismatches 55; Indels 38; Gaps 8;

QY 3 DSNOGNNOQYQOYSONG---NOQGN-----RYQGYQAYNMAQAPAGGYQNYQYSG 54  
Db 22 DQNOQYODPNOQOQFNQSGFDPNOQOQFNQSGFDPNOQYQODPNOQOQFNQAGFDPNOQYQOD 81  
QY 55 -----YQGGYQOYNPDAGY-QQOYNPGGYYQOYNPGGYYQOQF--NP 94  
Db 82 PNOQOQFNQSGFDPNOQYQODPNOQOQFN-QAGFDPNOQYQODPNOQOQFNQSGFDPNOQYQODP 140  
QY 95 QGGRGNKYNFNNNNNLOQYQAGFQ-----POSQ 122  
Db 141 NQOQFNQSPFDLNNQ-QFNQPGFNQSPAFETFTQEQ 175

RESULT 9  
FUS\_MOUSE STANDARD; PRT; 518 AA.

AC P56959;

DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE RNA-BINDING PROTEIN FUS (PIGGEN PROTEIN).

GN FUS.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN SEQUENCE FROM N.A.

RA Alapatt S.R., Zhang M., Zhao X., Allegro M.A., Allegro M.C.,

RA Burdall C.A.;

RT "Regulation of p1pen expression in mouse embryos";

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: BINDS BOTH SINGLE-STRANDED AND DOUBLE-STRANDED DNA AND

CC PROMOTES ATP-INDEPENDENT ANNEALING OF COMPLEMENTARY SINGLE-

CC STRANDED DNAs AND D-LOOP FORMATION IN SUPRHELICAL DOUBLE-STRANDED

CC DNA. MAY PLAY A ROLE IN MAINTENANCE OF GENOMIC INTEGRITY (BY

CC SIMILARITY).

CC -1- SUBUNIT: COMPONENT OF NUCLEAR RIBOPROTEIN COMPLEXES (BY

CC SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).

CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: AF224264; AAF70602.1; -

DR MGD: MGI:1353633; FUS.

DR InterPro: IPR002952; Eggshell.

DR InterPro: IPR000504; RRM.

DR InterPro: IPR001876; Znf-RanBP.

DR Pfam: PF00076; rrm; 1.

DR Pfam: PF00641; zf-RanBP; 1.

DR SMART: SM00360; RRM; 1.

DR SMART: SM00547; Znf-RBZ; 1.

DR PROSITE: PS0102; RRM; 1.

DR PROSITE: PS00030; RRM\_RNP\_1; FALSE\_NEG.

KW RNA-binding; DNA-binding; Nuclear protein; Repeat; Zinc-finger; Zinc;  
Metal-binding.

FT DOMAIN 1 167 GLN/GLY/SER/TYR-RICH.

FT DOMAIN 168 265 GLY-RICH.

FT DOMAIN 278 364 RNA-BINDING (RRM).

FT DOMAIN 364 518 ARG/GLY-RICH.

FT ZN\_FING 421 440 C4-TYPE (POTENTIAL).

SO SEQUENCE 518 AA; 52673 MW; E06F231BFED78D6 CRC64;

Query Match 20.4%; Score 143; DB 1; Length 518;  
Best Local Similarity 29.1%; Pred. No. 6.5e-05;  
Matches 46; Conservative 12; Mismatches 60; Indels 40; Gaps 7;

QY 2 SDSNOGN--NOQNYQOYSONGNOQGNRRQGYDAYNMAQAPAGGYQNYQYSGYSGYG- 58  
Db 85 SOSOSSSYGGQSSYPTGGQAPSPSTGSGYGGSSOSSSYGQPOSGGTYGGQSSGTYGGQSSY 144  
QY 59 GYQOQ--YNPAGY--QQOYNPQ-----GY-----QOYNPQ 85  
Db 145 GQOQSSSYNPQGYQOQNYSSSGGGGGGGGAGNYGQDQSSMSGGGGGGGGYGNQDQSSGGG 204  
QY 86 GGYOQOQFNPOGGRGNYKFNFNNNNLOQYQAGFQPOSOQ 123











```

OX NCBL_TaxID-10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Kidney;
RA Arao Y., Kikuchi A.;
RT "Differential expression of AUP1 isoforms in rat tissues.";
RL Submitted (JUL-2000) to the EMBL/Genbank/DDJ databases.
CC -1- FUNCTION: BINDS WITH HIGH AFFINITY TO RNA MOLECULES THAT CONTAIN
CC AU-RICH ELEMENTS (ARES) FOUND WITHIN THE 3-PRIME UNTRANSLATED
CC REGIONS OF MANY PROTOGENES AND CYTOKINE MRNAS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR; COMPONENT OF RIBONUCLEOSOMES (BY
CC SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS: 1/P45 (SHOWN HERE), 2/P42, 3/P40
CC AND 4/P37; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AB046615; BAB03465.1; -
DR EMBL: AB046616; BAB03466.1; -
DR EMBL: AB046617; BAB03467.1; -
DR EMBL: AB046618; BAB03468.1; -
DR InterPro: IPR000504; RRM.
DR Pfam: PF00076; Rrm; 2.
DR SMART: SM00360; RRM; 2.
DR PROSITE: PS0102; RRM; 2.
DR PROSITE: PS00030; RRM_RNP_1; 2.
KW Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat;
KW Alternative splicing.
FT DOMAIN 11 45 ALA-RICH.
FT DOMAIN 95 177 RNA-BINDING (RRM) 1.
FT DOMAIN 180 259 RNA-BINDING (RRM) 2.
FT DOMAIN 268 345 GLY-RICH.
FT DOMAIN 292 330 TYR-RICH.
FT VARSPLIC 77 95 MISSING (IN ISOFORM 2 AND ISOFORM 4).
FT VARSPLIC 283 332 GPSQNMNGYSNYMNGYSGYNSQYGGYGGYDYGYS
FT YGCGDYSN -> D (IN ISOFORM 3 AND ISOFORM
FT 4).
SO SEQUENCE 353 AA; 38192 MW; 6190EEL006F07DCB CRC64;

```

```

Query Match 18.3%; Score 128; DB 1; Length 353;
Best Local Similarity 38.5%; Pred. No. 0.00072;
Matches 30; Conservative 5; Mismatches 33; Indels 10; Gaps 1;

```

```

QY 17 SONGNQOQGNNRYQGYQATYAAQAPAGYQNTQGYSGYQGGYQOYNPDAGYQOQYNPQ 76
DB 285 SQMWNQGYSNYMWNGY-----GSYGYNSQGYGGYGGYDYGYSNYSGYDYSNQ 334
QY 77 GGYQATNPQGYQOQFNP 94
DB 335 SGYGYVSRNGHQNSTYKP 352

```

Search completed: February 1, 2002, 15:07:32  
Job time: 312 sec